UNIVERSITY OF LONDON THESIS

Degree PhD Year 2007 Name of Author SMITH, S. Lan-Lan

ELUCIDATING MOLECULAR MECHANISMS OF LEUKAEMOGENESIS IN NORMAL KARYOTYPE AML

being a thesis submitted to the

University of London

for the degree of

Doctor of Philosophy

by

S. Lan-Lan Smith

University College London

Cancer Research UK

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

DECLARATION OF WORK PRESENTED

I, S. Lan-Lan Smith, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

ACKNOWLEDGEMENTS

I would like to sincerely thank my supervisors Jude Fitzgibbon and Dominique Bonnet for giving me the opportunity to do my PhD in their labs and to all the guidance and support they have given me over the course of my studies. I have learnt so much over the past three years and continue to learn from them.

I would also like to acknowledge all the members of the Medical Oncology department at Charterhouse Square and members of the Bonnet lab. They have been a wonderful source of collaborations, friendships, corrections and support. I'd like to give a special thank you to Daniel Pearce, who helped me greatly during the course of my PhD and with the critical reading of this manuscript.

I would also like to thank my family and my partner for all their love and support. I will always count my years of doing my PhD as a special experience and without everyone's support and assistance I'm not sure I would have made it through. There are so many people who have made a significant positive impact over the past three years it would be impossible to list them all, but thank you everyone.

ABSTRACT

The aim of this thesis was to examine the role that gene specific mutations play in leukaemogenesis of normal karyotype (NK) acute myeloid leukaemias (AMLs), by studying the number of genes involved, types of mutations, order by which they arise and their function. AML is a stem cell disease in which the bulk of the disease is perpetuated by a rare population of leukaemic stem cells (LSCs), which arise due to an accumulation of genetic events within the target cell.

Eight genes known to be mutated in AML were screened from a panel of 88 NK AML patients. One hundred and twenty-seven mutations were detected; multiple mutations were common and 44% of the AMLs contained mutations that affect both differentiation and proliferation, indicating a requirement for cooperating mutations. Regions of acquired uniparental disomy (aUPD) detected by SNP arrays were found to contain homozygous mutations in genes implicated in AML, such as homozygous *FLT3*-ITDs on 13q, indicating that mitotic recombination can act as a second genetic event in leukaemogenesis. Semi-quantitative real-time PCR mutation specific assays, were designed against mutations in the transcription factor *CEBPA*. These assays were used to test AML samples which had been sorted into populations corresponding to stem cell and early progenitors in normal haematopoiesis. *CEBPA* mutation was detected in all compartments tested, indicating it is an early event in leukaemogenesis, which is supported by their occurrence in familial AML. Functional studies of mutations in *CEBPA* indicated that the 30kDa isoform of the protein was able to direct differentiation of lineage depleted cord blood mononuclear cells towards the myeloid pathway, where the addition of an ITD within the C-terminal region promoted erythroid differentiation. The work presented here demonstrates that gene specific mutations play a key role in leukaemogenesis.

PUBLICATIONS

Summers K, Stevens J, Kakkas I, Smith M, Smith LL, MacDougall F, Cavenagh J, Bonnet D, Young BD, Lister TA, Fitzgibbon J. (2007) Wilms tumour 1 (WT1) mutations are associated with FLT3-ITD and failure of standard induction chemotherapy in patients with normal karyotype AML. *Leukemia.* Mar; 21(3):550-1.

Smith LL, Pearce DJ, Smith ML, Jenner M, Lister TA, Bonnet D, Goff L and Fitzgibbon J. (2006) Development of a quantitative real-time PCR method for monitoring *CEBPA* mutations in normal karyotype acute myeloid leukaemia. *British Journal of Haematology,* Apr; 133(1):103-5.

Pearce DJ, Taussig D, Zibara K, Smith LL, Ridler CM, Preudhomme C, Young BD, Rohatiner AZ, Lister TA and Bonnet D. (2006) AML engraftment in the NOD/SCID assay reflects the outcome of AML: implications for our understanding of the heterogeneity of AML. *Blood,* Feb 1; 107(3):1166-73.

Fitzgibbon J, Smith LL, Raghavan M, Smith ML, Debernardi S, Skoulakis S, Lillington D, Lister TA and Young BD. (2005) Association between acquired uniparental disomy and homozygous gene mutation in acute myeloid leukemias. *Cancer Research,* Oct 15; 65(20):9152-4.

Smith ML, Arch R, Smith LL, Bainton N, Neat M, Taylor C, Bonnet D, Cavenagh JD, Andrew Lister T and Fitzgibbon J. (2005) Development of a human acute myeloid leukaemia screening panel and consequent identification of novel gene mutation in FLT3 and CCND3. *British Journal of Haematology,* Feb; 128(3):318-23.

INDIVIDUAL CONTRIBUTIONS

Chapter 3:

Mutation screening of *CCND3*: Lan-Lan Smith and Matt Smith

Mutation screening of *FES:* Lan-Lan Smith

Mutation screening of *NPM1* and *WT1*: Karin Summers and Lan-Lan Smith

Mutation screening of *FLT3, CEBPA, NRAS, KRAS, RUNXI:* Matt Smith

Chapter 4:

SNP arrays: Manoj Raghavan

Mutation screening of *FLT3, CEBPA, WT1, PU.1, CCND3, RUNX1:* Lan-Lan Smith

Design and testing of mutation specific primers for *CEBPA*: Lan-Lan Smith

Design and testing of real-time probes for *CEBPA*: Lan-Lan Smith

Sorting of patient AML samples: Lan-Lan Smith and Daniel Pearce

Real-time analysis of sorted AML samples: Lan-Lan Smith

Chapter 5:

Cloning of mutant *CEBPA*: Lan-Lan Smith

Testing of construct: Lan-Lan Smith

Production of lentivirus expressing *CEBPA* mutants: Lan-Lan Smith

Transduction of lineage depleted cells: Lan-Lan Smith

In vivo NOD/SCID experiment: Lan-Lan Smith and Daniel Pearce

In vitro methylcellulose experiments: Lan-Lan Smith

ABBREVIATIONS

2% FCS: PBS supplemented with 2% FCS

ALL: acute lymphoid leukaemia

AML: acute myeloid leukaemia

APC: allophycocyanin

APL: acute promylocytic leukaemia

APS: ammonium persulphate

ATRA: all-trans retinoic acid

BM: bone marrow

bp: base pair

BSA: bovine serum albumin

CD: cluster of differentiation

CFC: colony forming cell

CLL: chronic lymphoid leukaemia

CLP: common lymphoid progenitor

CML: chronic myeloid leukaemia

CMP: common myeloid precursor

CR: complete remission

CRUK: Cancer Research UK

dH_2O: de-ionised water

DAPI: 4',6-diamidino-2-phenylindole

DHPLC: denaturing high performance liquid chromatography

DMEM: Dulbecco's Modified Eagle's Medium

DMSO: dimethyl-sulphoxide

DNA: deoxyribonucleic acid

DTT: 1,4-Dithiothreitol

EDTA: ethylenediaminetetraacetic acid

FAB: French-American-British

FACS: Fluorescence-activated cell sorting

FCS: fetal calf serum

FISH: fluorescence in situ hybridization

FITC: fluorescein isothiocyanate

g: grams

g: gravity

GFP: green fluorescent protein

GMP: granulocyte/macrophage progenitor

HSC: haematopoietic stem cell

IMDM: Iscove's Modified Dulbecco's Medium

IRES: internal ribosome entry site

ITD: internal tandem duplication

JM: juxtamembrane

kDa: kilo Dalton

L: litre

Lin-: lineage depleted

LSC: Leukaemic stem cell

M: molar

MDS: myelodysplastic syndrome

MEP: Macrophage/Erythrocyte population

MNC: mononuclear

MPD: myeloproliferative disease

MRC: Medical Research Council

MRD: minimal residual disease

NK: normal karyotype

NOD/SCID: non obese diabetic/severe combined immunodeficiency

ORF: open reading frame

PAGE: polyacrylamide gel electrophoresis

PB: peripheral blood

PBS: phosphate buffered saline

PCR: polymerase chain reaction

PE: phycoerythrin

PEI: polyethylenimine

PerCP: peridinin-chlorophyll protein complex

RNA: ribonucleic acid

rpm: revolutions per minute

RPMI: Roswell Park Memorial Institute

SAP: shrimp alkaline phosphatase

SDS: sodium dodecyl sulfate

SOC: Super Optimal Broth + catobolite repression

SP: side population

SSCP: single-strand conformational polymorphism

TBS: Tris-buffered saline

TEAA: triethylammonium acetate

TEMED: N,N,N',N'-Tetramethylethylenediamine

TKD: tyrosine kinase domain

UPD: uniparental disomy

μL: microlitre

UK: United Kingdom

UTR: untranslated region

WHO: World Health Organization

GENE NAMES

β2M: Beta-2 microglobulin

ABL: v-abl Abelson murine leukemia viral oncogene homolog 1

AF9: ALL1 fused gene from chromosome 9

AML1: acute myeloid leukemia 1

BCR: breakpoint cluster region

Bmi1: murine leukaemia viral onocogene homologue

CBF: core binding factor

CCND3: cyclin D3

CDK: cyclin dependant kinase

CEBPA: CCAAT enhancer binding protein alpha

cKIT: Proto-oncogene tyrosine-protein kinase Kit

E2F: family of transcription factors

ENL: eleven-nineteen leukaemia

ERK: extracellular signal-regulated kinases

ETO: Eight twenty-one

ETS1: v-ets erythroblastosis virus E26 oncogene homolog 1

EVI1: Ecotropic viral integration site 1

GAS7: growth arrest-specific 7

GATA-1: globulin transcription factor 1

G-CSF: granulocyte-colony stimulating factor

FES: Feline sarcoma proto-oncogene

FLT3: fms-like tyrosine kinase 3

MDS1: myelodysplasia syndrome 1

MEK: MAPK/ERK kinase

MLL: mixed lineage leukaemia

MYH11: myosin heavy chain 11

NPM: nucleophosmin

PDGFR: platelet-derived growth factor receptor

PML: promyelocytic leukemia

PU.1: PU-box 1

RARA: retinoic acid receptor alpha

RAS: rat sarcoma 2 viral oncogene homolog

Rb: retinoblastoma

RUNX1: runt-related transcription factor 1

SCL: stem cell leukaemia

STAT: Signal transducer and activator of transcription

Tal-1: T-cell acute lymphocytic leukaemia 1

TEL: ETS variant gene 6

TP53: tumor protein 53

TRIB1: tribbles homologue 2

UNG: uracil-DNA glycosylase

WT1: Wilms' tumour suppressor gene

Amino Acid and DNA Base Abbreviations

Amino Acids

A Ala alanine

C Cys cystine

D Asp aspartatic acid

E Glu glutamic acid

F Phe phenylalanine

G Gly glycine

H His histidine

I Ile isoleucine

K Lys lysine

L Leu leucine

M Met methionine

N Asn asparagine

P Pro proline

Q Gln glutamine

R Arg arginine

S Ser serine

T Thr threonine

V Val valine

W Trp tryptophan

Y Tyr tyrosine

DNA/RNA

A adenine

C cytosine

G guanine

T tryosine

U uracil

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

BLOOD.

A word that evokes many images, blood is a substance that binds human life together. Over the years blood has come to represent many things, from family ties to hard work. Blood has been called upon by literary heroes like Shakespeare to describe courage and has even come to be a euphemism for life itself. When people die in violent ways, most often at the hands of fellow man, blood is spilt.

In many belief systems around the world and throughout history, blood plays a central role. The ancient Aztec civilization believed that the future could be prophesised by blood sacrifice of enemies and the spilling of blood from their high priests. Blood sacrifice was also practiced by ancient Germanic tribes, the sacrifice often of animals but human blood letting was also involved. The act of sprinkling blood associated with these rituals was called *bleodsian* in Old English and subsequently gave rise to a word that is very familiar today, *blessing*. To Catholics the ritual of communion remains central to their beliefs in which they partake of Christ's body and blood. Even in secular rituals blood plays a central role. The "blooding," occurs in fox hunting, where the lead huntsman smears the blood of the fox onto the cheeks of a newly initiated hunt follower. In a more light-hearted manner, a nosebleed in Japanese culture has come to represent sexual frustration or desire.

Although blood coveted such importance in what it came to represent, it itself as a substance has been poorly understood for the majority of human history. For thousands of years the human body and its inner workings were a mystery; illnesses were treated with a combination of trial and error and

mysticism, of which such efforts were met with varying degrees of success. Hippocrates considered blood to be one of the four humours that made up the body and that illness was a result of imbalance between these humours. Evidence of blood-letting in practice dates as far back as 2500 BCE by the Egyptians and the technique remained in vogue until well into the 19th century.

As time passed the study of anatomy and physiology gradually stripped away the layers of mystery that encompassed the human body, but understanding of blood remained elusive until the advent of microscopy. Human red blood cells were first described in detail by the pioneering microscopist Anton van Leeuwenhoek in 1674, whereas the white blood cells or "globuli albicantes" as they were called were first noted by Joseph Lieutaud in 1749.

It comes of no surprise that the description and understanding of blood disorders is of relatively recent origin as well. Prior to microscopy, most afflictions of the blood, such as anaemia and some infectious diseases, were recognized not by studying the blood itself, but by associated enlargement of the spleen. Although a small number of cases of patients with uncommon or peculiar alterations of the blood were described early in the 19th century, it was not until the seminal papers by doctors John Bennett and Rudolf Virchow were published in 1845 that leukaemia was characterised as a disease in its own right.

Since those early days, understanding of leukaemia has leapt forward drastically and research into the disease has led the field of modern cancer research. Developments in microscopy, cellular staining and cytochemical techniques have resulted in better characterisation of leukaemias, in particular

recognition of its heterogeneous nature. Within the past twenty years the development of monoclonal antibodies and the recognition of patterns of cell surface antigen expression have allowed scientists to further categorise different cell types that make up leukaemia.

As molecular techniques have improved over the years, research into leukaemia has also shifted gears, turning emphasis from phenotype to genotype. The discovery of the Philadelphia chromosome in chronic myeloid leukaemia was the first evidence that chromosomal damage is associated with cancer. Mustard gases developed for use in the Second World War marked the birth of chemotherapy as they went from being an agent of harm to helping treat leukaemia. Leukaemia is also the first cancer to be considered a "stem cell" disease, where a rare population of leukaemic stem cells are able to propagate the bulk of the disease.

A simple definition of leukaemia is a 'group of disorders characterised by the accumulation of malignant white blood cells in the bone marrow'. However with study, it soon becomes clear that leukaemia is no random mess of malignant cells, but there is order and a distinctive hierarchy involved. As Dante strove to understand hell by stratifying it into discriminate layers, scientists today strive to do the same with the process of leukaemogenesis.

CHAPTER 1: INTRODUCTION

OVERVIEW

Acute myeloid leukaemia (AML) is a clonal, malignant proliferation of undifferentiated haematopoietic cells that is morphologically, genetically and therapeutically heterogeneous. In normal haematopoiesis, mature cell types arise from the pluripotent haematopoietic stem cell (HSC). In a similar fashion, the leukaemic blast population, which has a low proliferation potential, is perpetuated by a small population of leukaemic stem cells (LSCs) (**Figure 1.1**). As such, the response of the LSC to therapy is thought to be essential in determining the patient's outcome, and in a large part reflect the genetic damage within these cells[1].

The relative rarity of LSCs compared to the myeloblasts which compose the bulk of the disease complicates the treatment of AML (**Figure 1.2**). The first phase of treatment, termed "induction therapy", attempts to destroy as many of the leukaemic blasts as possible in order to achieve complete remission (CR). CR is defined as the presence of less than 5% blasts in the bone marrow after treatment. Approximately 50-75% of *de novo* adult AMLs are able to achieve CR with induction therapy, though this is dependent on a number of prognostic factors. Once in remission, "consolidation therapy" is given in order to kill any remaining blast cells[2].

Unfortunately, despite good rates of CR in AML, many patients subsequently relapse from their disease and die. The majority of relapses

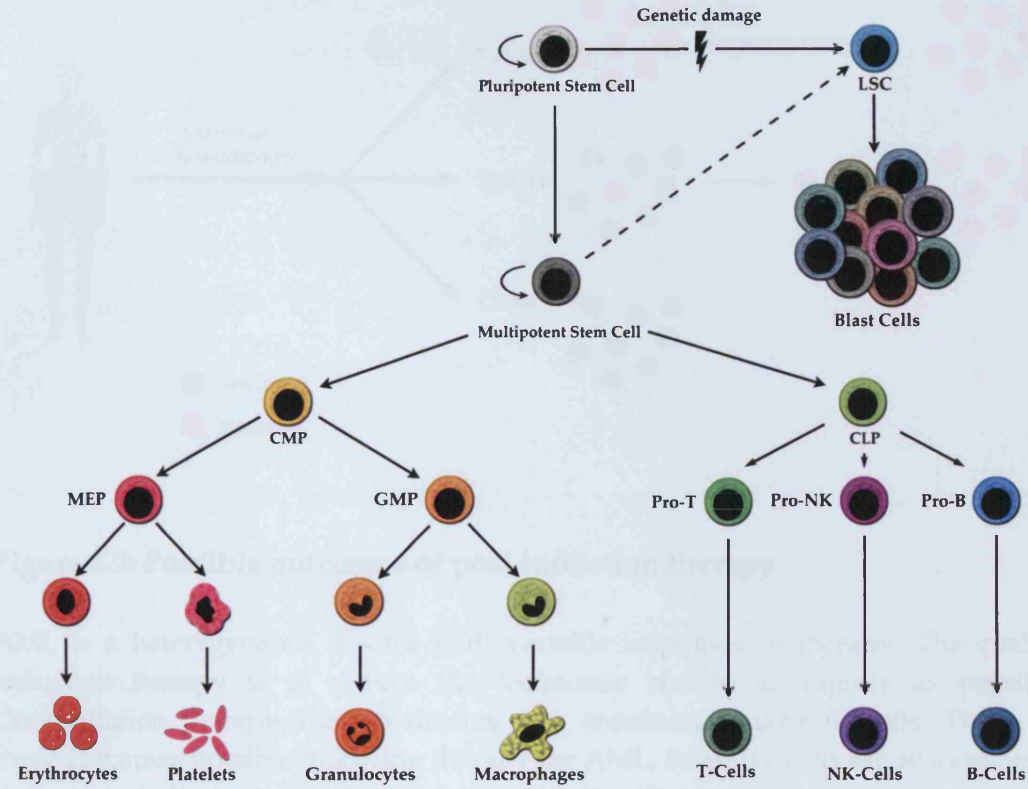


Figure 1.1: Normal and aberrant haematopoiesis

Normal haematopoiesis is driven by the haematopoietic stem cell (HSC), a pluripotent stem cell, which is able to self-renew and differentiate into all mature blood cell types. In AML, proliferation is dysregulated and differentiation is blocked, resulting in a mass of undifferentiated, non-functional blast cells perpetuated by the leukaemic stem cell (LSC). The LSC arises from an accumulation of genetic damage that targets genes involved in regulating proliferation and differentiation within HSCs or early progenitor cells. (Adapted from [3])

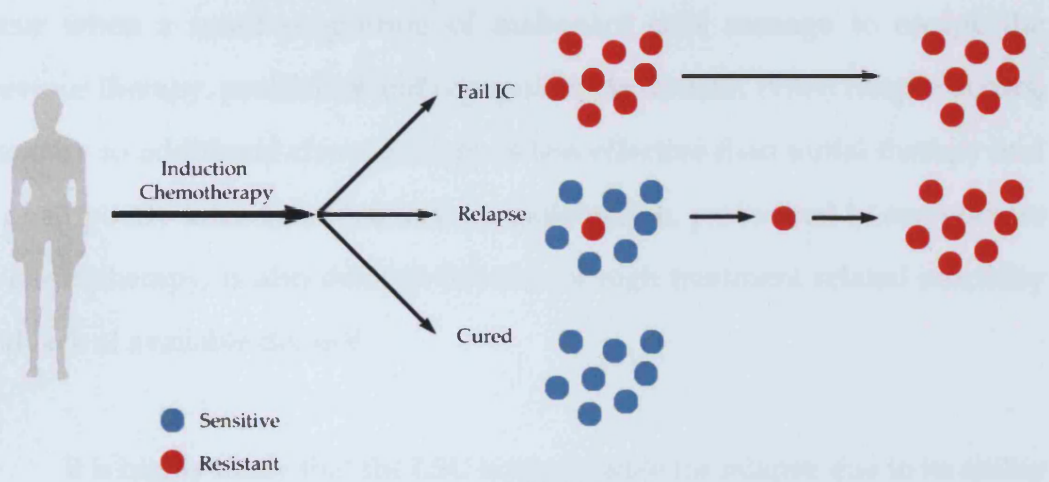


Figure 1.2: Possible outcomes of post induction therapy

AML is a heterogeneous disease with variable responses to therapy. The goal of induction therapy is to reduce the leukaemic burden as rapidly as possible. Consolidation therapy aims to destroy any remaining leukaemic cells. There are three outcomes possible following therapy for AML. Sensitive cells are susceptible to therapy and eliminated. If the majority of blast cells are resistant, therapy fails and the disease persists. Often the bulk of the disease is eliminated and a remission is achieved, with the bone marrow consisting of less than 5% blast cells. However cells resistant to therapy persist and eventually repopulate the disease, resulting in relapse of the disease.

occur when a small proportion of malignant cells manage to escape the previous therapy, proliferate and repopulate the disease. When relapse occurs, response to additional chemotherapy is less effective than initial therapy and is often poorly tolerated. Stem cell transplantation, performed in conjuncture to chemotherapy, is also difficult because of high treatment related mortality and lack of available donors[2].

It is highly likely that the LSC is responsible for relapse due to its ability to give rise to leukaemic blasts and its largely quiescent nature, which allows it to escape conventional chemotherapy[4]. An understanding of the origin of the LSC is essential to elucidate the mechanisms of AML development and progression.

The LSC arises due to an accumulation of genetic events within the HSC or early progenitors. However, examination of the disease is focused on its progeny, the leukaemic blasts. Cytogenetic analysis of the blast cells, which reveals chromosomal aberrations, can be used to predict response to treatment (the patient's prognosis) however approximately fifty percent of AML cases have normal cytogenetics. Therefore, it is necessary to look beyond the resolution of cytogenetics to the molecular level and to examine gene specific mutations[5]. In AML, analysis of mutation status has been useful for risk stratification of patients, disease monitoring and the development of novel therapies.

Although knowledge of mutation status is useful for patient therapy; to more fully understand the disease it is important to address the question of how the LSC arises. As it is genetic aberrations which dictate the nature of the disease, there are several questions to address:

What types of mutations are present?

How many genes are involved?

What length of time does it take for the disease to develop?

In which order do the mutations arise?

What is the function of mutations in AML?

In this thesis we aim to address these questions. Mutation screening of a panel of normal karyotype (NK) AML patients allows us to see the number and pattern of gene mutations involved in leukaemogenesis. Long latency periods observed in familial AMLs indicate several genetic events are required for the development of overt AML. The discovery of acquired uniparental disomy giving rise to homozygous gene mutations shows clear progression of genetic events over time. To study the function of mutations in AML, we have examined the affects of mutations in the transcription factor *CEBPA* on normal haematopoiesis.

ACUTE MYELOID LEUKAEMIA

INCIDENCE

Leukaemia accounts for approximately 3% of all cancers in the United Kingdom. Grouped together, leukaemias are the ninth most common cancer in men and tenth in women. Age is an important risk factor in the development of AML. Although AML comprises roughly 75% of all cases of acute leukaemia it accounts for only 25 – 30% of cases of acute leukaemia in children aged 10 – 15 and less than 15% of cases of acute leukaemia in children under age 10. The incidence of AML rises dramatically for people over the age of 50 (**Figure 1.3**) [6,7]. Unfortunately, older adults respond poorly to current therapies and most patients with AML will die from the disease.

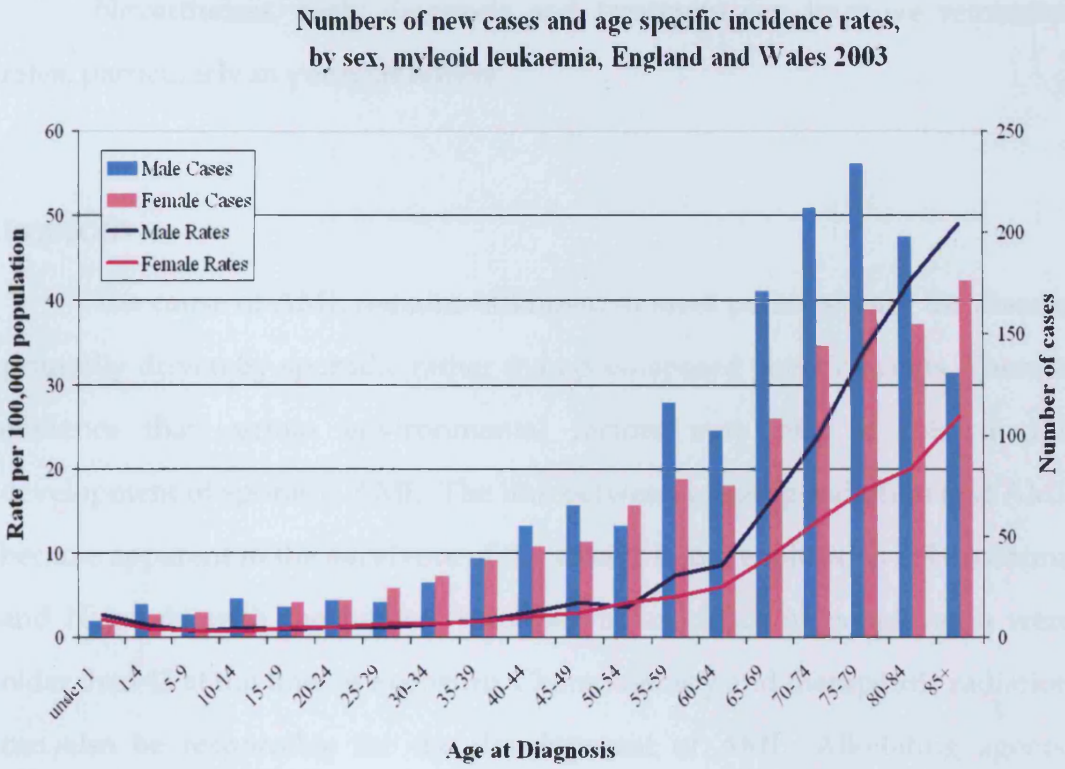


Figure 1.3 Rates and Incidence of myeloid leukaemias in England and Wales, 2003[7]

As age increases the rates and incidence of leukaemia also increase. Leukaemia is a disease that primarily affects older individuals, with age being a substantial risk factor for development of the disease. Leukaemia is the 9th and 10th most common cancer in men and women respectively.

Nevertheless, early diagnosis and treatment can improve remission rates, particularly in younger adults.

ETIOLOGY

The cause of AML remains unknown in most patients; with the disease primarily driven by sporadic rather than predisposed genetic events. There is evidence that certain environmental factors may play a role in the development of sporadic AML. The link between ionizing radiation and AML became apparent in the survivors of the atomic bomb explosions at Hiroshima and Nagasaki with the highest increases in incidence in people who were older than 45 at the time of exposure. Chemotherapy and therapeutic radiation can also be responsible for the development of AML. Alkylating agents, etoposides and topoisomerase II inhibitors have all been implicated in the development of secondary AML, following the treatment of other cancers. Significant exposure to chemicals such as benzene and other petrolum derivatives has also been associated with an increased risk[2, 8].

Certain inherited genetic conditions are also associated with an increased risk for the development of AML. Incidence of AML in Down's syndrome (trisomy 21) is higher than that of the general population. Also associated with an increased risk of AML are Fanconi's anemia and Bloom's syndrome, both of which are characterised by chromosome fragility[2]. Familial cases of syndromes such as neurofibromatosis and Noonan Syndrome have been useful for discovery of mutated genes that also yield leukaemia if they arise sporadically.

Clinical Presentation and Diagnosis

Pathologically, the acute leukaemias are defined as malignancies of immature haematopoietic cells. Blast cells constitute greater than 20% of the bone marrow cells and cause failure to produce normal cells resulting in cytopenias. The clinical presentation is usually a manifestation of these cytopenias and unrelated to the blast cells themselves. While clinical findings vary for person to person, common symptoms include anemia, bruising and haemorrhages.

Anaemia, while common, is not a universal symptom of AML, presumably because of rapid onset of the disease in some cases. Neutropenia and thromobocytopenia are usually present along with a reduction in numbers of normal leucocytes. Bleeding as the result of a diminished platelet count is another common finding, and can manifest itself as petechial hemorrhages observable on the skin or through gross haemorrhage[8].

Organ infiltration by leukaemic blasts can be another presentation of leukaemia. The bone marrow is consistently infiltrated, but this cannot be observed upon physical examination. Enlargement of the gums, spleen and liver are rare but clinically assessable signs of leukaemic infiltration. Ultimately, leukaemia must be diagnosised by examination of the bone marrow itself[8].

Classification

Examination of the peripheral blood can often provide adequate information to diagnose leukaemia. Careful morphological assessment of normal and abnormal cell components of the peripheral blood and bone marrow remains essential to the diagnosis of the disease. Traditionally, AMLs

were classified into different phenotypic groups based on the French-American-British (FAB) Cooperative Group classification scheme, which split the disease into subtypes based on differentiation (**Table 1.1**). This scheme was recently replaced by the World Health Organization (WHO) classification system.

The discovery of the Philadelphia chromosome in 1960 heralded a new era in cancer genetics. The realisation that this small marker chromosome was the result of a chromosomal translocation began a period of intensive study into leukaemic karyotypes, revealing many more recurrent chromosomal abnormalities[9]. Cytogenetics have since been used in prognosis, allowing the segregation of AML into potential outcome groups, either as favourable, intermediate, or poor (**Table 1.2, Figure 1.4**)[10]. AML with normal karyotype is classified within the intermediate prognosis group, however research has recently shown that the presence of certain molecularly detectable mutations can be used to further risk stratify[5], which will be discussed in greater detail later in this thesis. The current classification system for malignant haematological conditions developed by the WHO takes into account cytogenetics. Leukaemias with consistent cytogenetic abnormalities and those which are related to myeloproliferative disorders (MDS) are divided into separate groups with the remaining following the traditional FAB system of classification (**Table 1.3**)[11].

THERAPY

Despite advances in the therapy of AML, the majority of patients will die from their disease. As discussed above, the most important prognostic factors determining the outcome of therapy are the acquired genetic changes

FAB Subtype	Name	Approximate % of adult AML patients
M0	Undifferentiated AML	5%
M1	Myeloblastic leukaemia with minimal maturation	15%
M2	Myeloblastic leukaemia with maturation	25%
M3	Promyelocytic leukaemia	10%
M4	Myelomonocytic leukaemia	20%
M4 eos	Myelomonocytic leukaemia with eosinophilia	5%
M5	Monocytic leukaemia	10%
M6	Erythroid leukaemia	5%
M7	Megakaryoblastic leukaemia	5%

Table 1.1: FAB classification of AML

The French-American-British (FAB) classification system divides AML into 8 subtypes, M0 through to M7, based on the type of cell from which the leukaemia develops and its degree of maturity. Diagnosis is made based on phenotype of the cells. Different FAB types tend to have different prognosis outcomes, although this is based on the underlying geneotype of the disease.

Risk Group	Abnormality	5 year survival	Relapse Rate
Favourable	t(8;21) t(15;17) inv(16)	70%	33%
Intermediate	Normal +8 +21 +22 del(7q) del(9q) Abnormal 11q23 All other structural/ numerical abnormalities	48%	50%
Adverse	-5 -7 del(5q) Abnormal 3q Complex	15%	78%

Table 1.2: Cytogenetic risk groups of AML

Hierarchical prognostic classification, derived taking into consideration the influence of additional cytogenetic abnormalities on outcome, and used for directing treatment approach in the current MRC AML 12 trial. The presence of certain genetic abnormalities indicates whether the AML will respond favourably or poorly to therapy. Prognosis may also be used to determine the nature of the therapy given, with more aggressive treatment for disease with adverse cytogenetics. (Adapted from [10])

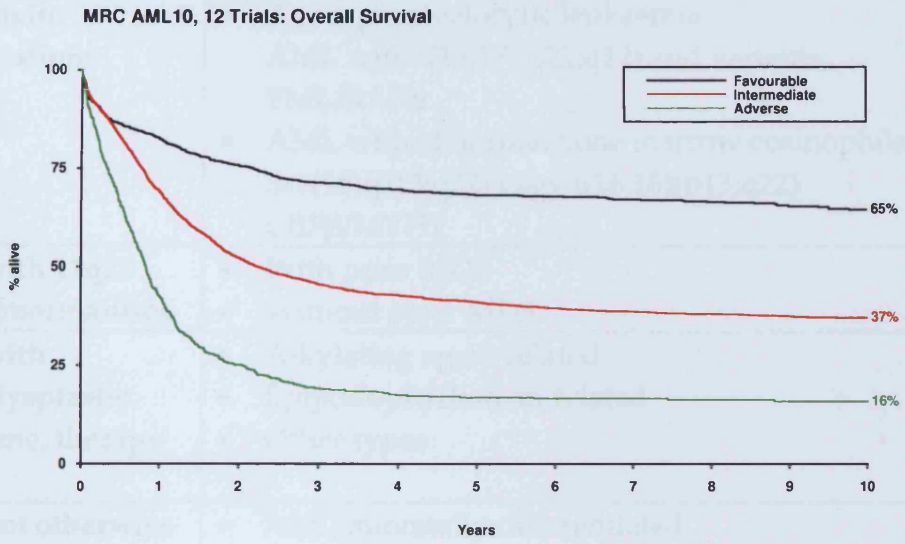


Figure 1.4: Prognostic risk based on karyotype in de novo AML

Overall survival of patients with AML is highly influenced by karyotype. In the MRC AML 10 and 12 trials, at 10 years, 65% of patients with favourable cytogenetics remained alive, 37% and 16% with intermediate and adverse cytogenetics survived for the same period of time respectively. (Adapted from [10])

AML with recurrent cytogenetic translocations	• AML with t(8;21)(q22;q22) AML1/CBFα/ETO • Acute promyelocytic leukaemia: AML with t(15;17)(q22;q12) and variants PML/RARα • AML with abnormal bone marrow eosinophils inv(16)(p13;q22) vagy t(16;16)(p13;q22) CBFβ/MYH1
AML with 11q23 MLL abnormalities	• With prior MDS • Without prior MDS
AML with myelodysplastic syndrome, therapy related	• Alkylating agent related • Epipodophyllotoxin related • Other types
AML not otherwise categorised	• AML minimally differentiated • AML without maturation • AML with maturation • Acute myelomonocytic leukaemia • Acute monocytic leukaemia • Acute erythroid leukaemia • Acute megakaryocytic leukaemia • Acute basophilic leukaemia • Acute panmyelosis with myelofibrosis

Table 1.3: WHO classification of AML[11]

The current classification system for AML, developed by the WHO, takes into account cytogenetics and attempts to provide more prognostic information about the disease.

within the leukaemic cells. These are also used to dictate the therapy that an individual patient recieves.

Younger patients (< 60 years old) undergoing a standard course of induction therapy can achieve an initial CR rate of 60-75%. Patients with APL (AML FAB type M3) through the addition of all-trans retinoic acid (ATRA) are able to achieve CR rates up to 80%[12]. The choice of post-remission therapy is determined by prognostic risk factors. Patients with favourable cytogenetics have an initial response to induction of approximately 85% and with intensive postremission therapy, the overall survival at 5 years exceeds 50%. For intermediate risk cytogenetics, many patients undergo allogeneic haematopoietic stem cell transplantation if there is a suitable donor, while those without receive intensive post-remission chemotherapy. Patients with unfavourable cytogenetics have the poorest outcome. While initial response rates may exceed 50%, the overall long term survival remains poor, regardless of the nature of postremission therapy[13]. Additionally, the development of resistance to chemotherapy is of major concern, with subsequent therapies being less effective in the case of relapsed leukaemias.

Older patients, which make up the majority of AML cases, have a dismal long-term prognosis. These patients tend to have more unfavourable prognostic factors at presentation and their treatment is made more difficult by an inability to tolerate intensive chemotherapy[14]. Therapy for this group provides the greatest challenge to AML care and the development of less severe and more targeted therapies is thought to be essential. These targeted therapies would not only benefit the older poor prognostic group but younger patients as well[15].

MINIMAL RESIDUAL DISEASE (MRD) MONITORING

As many patients will relapse the possibility of predicting an impending relapse becomes imperative for deciding further therapeutic options. The persistence of malignant cells at a level below morphological detection is termed minimal residual disease (MRD). The molecular detection of MRD may identify patients at a higher risk of relapse. Due to the ability of these remaining cells to repopulate the blast population, it is not unreasonable to assume that in the majority of relapse cases, it is the LSC that escapes therapy and serves as the source for future relapsed disease.

The ability to detect low numbers of residual leukaemic cells would have a variety of benefits. It would allow for a more precise evaluation of the effectiveness of treatment, for MRD-adapted post-remission therapies and for earlier prediction of impending relapses prior to clinical manifestations[16].

The study of MRD relies on modern technical methodology to identify disease below morphologically detectable levels. If the original leukaemic cell carries a molecular or antigenic marker that distinguishes it from non-leukaemic cells, then all cells that arise from that clone should contain this same marker. This property would allow for the application of techniques such as real-time PCR or flow cytometry to detect or quantify residual leukaemic cells[17]. Sensitive methods include cell cytometry, PCR against specific genetic lesions and increasingly against gene specific mutations and fluorescence in situ hybridisation (FISH), all of which have relative advantages, disadvantages and sensitivity for detecting MRD, which are listed in **Table 1.4**[16].

Technique	Sensitivity	Disadvantage
Cytogenetics	1 : 20	Laborious
FISH	$10^{-1} - 10^{-3}$	Limited probes available, variety of breakpoints
PCR	Up to 10^{-6}	Technically demanding, susceptible to false positives and contamination
RT-PCR	Up to 10^{-6}	Complex, labour intensive
Flow cytometry	$10^{-1} - 10^{-4}$	Dependant on markers on leukaemia

Table 1.4: Methods of detection of MRD in AML

The persistence of these malignant cells at a level below morphological detection is termed minimal residual disease (MRD). The molecular detection of MRD may identify patients at a higher risk of relapse. (Adapted from [16])

Chromosomal translocations are a common target for PCR based MRD detection techniques; however this method of detection is limited due to a high proportion of normal karyotype AMLs. Gene expression as detected by real-time PCR also plays a key role in MRD monitoring. High expression of the Wilms' tumour suppressor gene (*WT1*) in AML during remission is associated with impending relapse of the disease[18]. Flow cytometric detection of aberrant immunophenotypes is another method utilised for MRD monitoring. The identification of new markers of leukaemia by gene array technology should lead to the design of simple and reliable antibody panels for universal monitoring of MRD[19]. Irrespective of the method of detection correlation between high MRD and relapse in patients indicates the need for development of increasingly sensitive and diverse techniques for the monitoring of disease.

THE HAEMATOPOIETIC SYSTEM

Since AML is the result of malignant haematopoiesis, the study of normal haematopoiesis and the processes by which it becomes deregulated is essential for understanding the disease.

NORMAL HAEMATOPOIESIS

The variety of cells in normal blood is constantly being maintained by the haematopoietic system, which produces all cells of the blood lineage. The turnover of cells in the haematopoietic system in an average human weighing 70kg is estimated to be close to one trillion cells per day, including 200 billion erythocytes and 70 billion neutrophilic leukocytes. This remarkable turnover is maintained by a small population of HSCs, which has one of the greatest powers of renewal of adult tissue[20]. Other components of the bone marrow include stromal cells, stromal stem cells, haematopoietic progenitor cells,

mature and maturing white, red blood cells and platelets and their precursor megakaryocytes.

HAEMATOPOETIC STEM CELLS

The HSC has the ability to both maintain itself and to differentiate into all cells of the blood lineage (**Figure 1.5**). A hierarchy exists within the system in which the pluripotent HSC has the greatest self-renewal and differentiation capacity of any haematopoietic cell. The hallmark properties of the HSC are the ability to balance self-renewal versus differentiation cell fate decisions to provide sufficient primitive cells to sustain haematopoiesis, while generating more mature cells with specialised capacities. Through the process of asymmetric cell division, a single division can result in the formation of both an identical stem cell and a more mature cell[21]. In order to assure a persistent pool of regenerating cells without outgrowth of immature cell types, tight regulation of HSC division is required. Unchecked growth of immature cells is thought to represent a paradigm for malignant outgrowth, at least for AML[3].

As cells become more differentiated, they lose the capacity for self-renewal. The HSC yields committed haematopoietic progenitors which are restricted in their developmental potential. The common myeloid progenitor (CMP) produces red blood cells (erythrocytes) and platelets as well as the myeloid series; the monocytes, neutrophils, eosinophils and basophils. The common lymphoid progenitor (CLP) gives rise to cells of the adaptive immune system, the B, T and natural killer cells[9].

HSCs have the potential to perform a variety of functions. They can self-renew, differentiate into a variety of specialised cells, mobilise out of the bone marrow into circulating blood, undergo apoptosis or remain in a

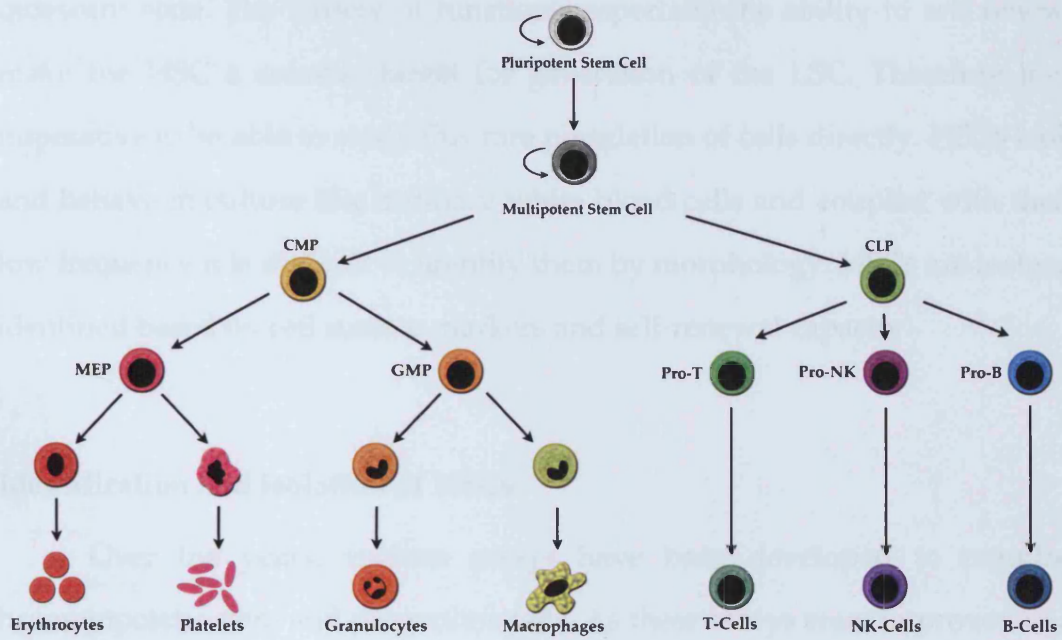


Figure 1.5: Normal haematopoiesis

The variety of cells in normal blood is constantly being maintained by the haematopoietic system, a hierarchy of cells which is driven by the pluripotent haematopoietic stem cell which has the ability of self renew and is able to produces all mature cells of the blood lineage. Common myeloid and lymphoid progenitors have limited self renewal capacity and are restricted to differentiate into their specific lineages. As cells become increasingly differentiated, their ability to self-renew decreases as well.

Mature blood cells are broadly split into three major groups. The erythrocytes are the oxygen carrying red blood cells, lymphoid cells are the cornerstone of the adaptive immune system, primarily consisting of T-cells and B-cells. The third lineage is the myeloid lineage which includes the granulocytes, megakaryocytes and macrophages. The myeloid lineage is involved in diverse roles such as innate and adaptive immunity and blood clotting. (Adapted from [3])

quiescent state. The variety of functions, especially the ability to self renew, make the HSC a suitable target for generation of the LSC. Therefore it is imperative to be able to study this rare population of cells directly. HSCs look and behave in culture like ordinary white blood cells and coupled with their low frequency it is difficult to identify them by morphology. HSCs are instead identified based on cell surface markers and self-renewal capacity.

Identification and isolation of HSCs

Over the years, various assays have been developed to examine haematopoietic stem and progenitor cells. As these assays were improved and progressively became more selective for stem cells, our definition of stem cells and understanding of haematopoiesis has developed.

The colony formation cell (CFC) assay involves the use of a semi-solid, methylcellulose-based medium supplemented with cytokines that promote differentiation and proliferation of haematopoietic progenitor cells. Long-term cultures of haematopoietic cells with stromal feeder layers are an attempt to recreate haematopoiesis *in vitro*. Within this system, committed cells and progenitors exhaust their proliferative potential within a few weeks and are no longer part of the culture, whereas cells with more extensive proliferative ability are able to form colonies beneath the stromal layer[22, 23]. This system is unable to assess for lymphoid and therefore, multilineage, potential and the assessment period is not excessively long. These problems were overcome by the development of systems which involved initially growing cells in a lymphoid supportive environment and subsequent transplant to a myeloid supportive environment[24,25]. These systems however remain limited by requirements for external stimulation by cytokines and therefore do not replicate endogenous conditions well.

Observing growth of cells in an endogeneous microenvironment where they would receive the appropriate signals would aid in overcoming these limitations. Xenotransplantation models involving inbred mice were developed which allowed the transplantation of haematopoietic cells from genetically similar mice to observe the growth of primitive haematopoietic for a significant proportion of time[26,27].

Immuno-compromised mouse models offer the most suitable environment for the study of human haematopoietic cells. The severe combined immunodeficiency (SCID) mutation in mice causes a lack of functional T and B lymphocytes and with cytokine stimulation, human cells may develop in the marrows of these mice[28]. When the SCID mutation is combined with the non-obese diabetic (NOD) mutation, mice also have an impaired natural killer and antigen-presenting cell function and the growth of human cells in these mice is possible without cytokine stimulation[29-31]. Given ideal conditions, the NOD/SCID model can support multilineage (B-cell and myeloid) engraftment for over 3 months and can also support secondary engraftment[32]. In this way, the NOD/SCID assay can assess the self-renewal, proliferation and differentiation of human haematopoietic stem cells.

The mouse model has revolutionised the field by allowing further progress to be made into isolating the HSC. Primitive haematopoietic cells, including the HSCs, do not express certain cell surface markers that are associated with mature blood cell types. Lack of expression of a variety of mature lineage (Lin) markers, can therefore be used to distinguish immature cells from the more abundant differentiated cells[33].

The expression of CD34 is commonly used to define the population of

HSCs and early progenitors. CD34 expression is observed on the majority of the cells with *in vivo* repopulating potential, including the HSCs and long term repopulating cells as well as early progenitors, but not the majority of terminally differentiated cells. HSCs can further be defined by their patterns of expression of other markers. The majority of Lin-/CD34+ cells are also CD38+; these cells have short-term self-renewal properties and are able to yield a transient wave of repopulation in immunodeficient mice. Lin-/CD34+/CD38- cells have the ability for sustained multilineage haematopoietic reconstitution after transplantation into immunodeficient mice. This subset is generally considered to contain the HSC, although more recent research has demonstrated that cells with long-term repopulating ability may additionally reside within a Lin-/CD34- population[33]. The further development of more permissive murine models should allow for further refinement of the identity of human HSCs.

Extensive work has gone into further refinement and characterisation of specific markers on the surface of human HSCs, however that is beyond the scope of this thesis to discuss.

HSC Fate Choice

A HSC has the ability to self-renew, differentiate, undergo apoptosis, or to remain in a quiescent state. The decision to follow one of these routes is ultimately determined at the level of differential gene activity within the cell. Gaining understanding of how these processes work is important to the study of AML and other haematological malignancies as the dysregulation between self-renewal and differentiation in HSCs is an essential feature of leukaemogenesis.

This sets out important questions, such as what types of mutations are required for this dysregulation and are there patterns of mutations commonly observed in AML.

LEUKAEMIA IS A STEM CELL DISEASE

As described in the overview, AML is a stem cell disease where the bulk of the disease, the leukaemic blasts, is perpetuated by the LSC. LSCs exhibit a dramatic increase in self-renewal ability over normal HSCs, and a block in differentiation. The presence of a long-term initiating cell in leukaemia with the ability to self-renew *in vivo* was demonstrated by xenotransplant models into immune deficient mice[34,35]. Leukaemia is not the only cancer which has been shown to have a stem cell origin, this has also been observed in brain, breast and more recently, colorectal cancers[36,37,38]. Furthermore, it has been demonstrated that the majority of LSCs exist in a quiescent state[39], thereby allowing them to escape therapies which are directed towards rapidly dividing cells. Leukaemia remains a useful system to study the process by which cancer stem cells arise due to good characterisation of the haematopoietic system and the HSC, and the ease of access to the cells.

SITE OF ORIGIN OF THE LSC

A hallmark of all cancers is the capacity for unlimited self-renewal which is also a characteristic of normal stem cells. Due to this shared attribute, it has been proposed that leukaemia arises from transforming events within the HSC. Furthermore, if multiple mutations are required for transformation to a malignant state, stem cells which have greater longevity compared to other cell types, have sufficient opportunity to acquire the requisite genetic damage. Alternatively, it is possible that the LSC arises from a more restricted

progenitor or even a differentiated mature cell which would first need to reacquire the ability to self renew before becoming tumourigenic (**Figure 1.6**)[40].

It is therefore likely that AML represents aberrant haematopoiesis with the target cell for transformation lying within the HSC repertoire. The normal HSC compartment is made up of pluripotent stem cells and progenitors with more restricted self renewal and differentiation ability. Similarly, LSCs are not functionally homogeneous but comprised of distinct hierarchically arranged LSC classes[41]. Therefore characterising the LSC compartments, and examining how mutations affect these compartments is also important to understanding AML. For most AML subtypes, the only cells which are able to repopulate the disease within the NOD/SCID model have a CD34+/CD38- phenotype, similar to the normal HSC, whereas the CD34+/CD38+ leukaemic blasts cannot transfer the disease to the mice[42] [35]. This however does not completely clarify the cell of origin which becomes a LSC.

Although the pluripotent HSC is an attractive target for transformation, there is experimental evidence which indicates that other types of early progenitor cells may serve as targets as well[40]. Cell populations transduced with a retroviral vector encoding the fusion proteins involving the mixed lineage leukaemia (*MLL*) gene could generate AML *in vivo* in a murine model of the disease. The fusion *MLL-ENL* was able to generate AML from transduction of HSCs, CMPs and GMPs. However it was not possible for erthyroid progenitors (MEP) to recapitulate AML. Similar studies using *MLL-GAS7*[43] and *MLL-AF9*[44] [45] fusion proteins demonstrated that some early progenitor populations may give rise to AML when transformed with these fusion proteins (**Figure 1.7**).

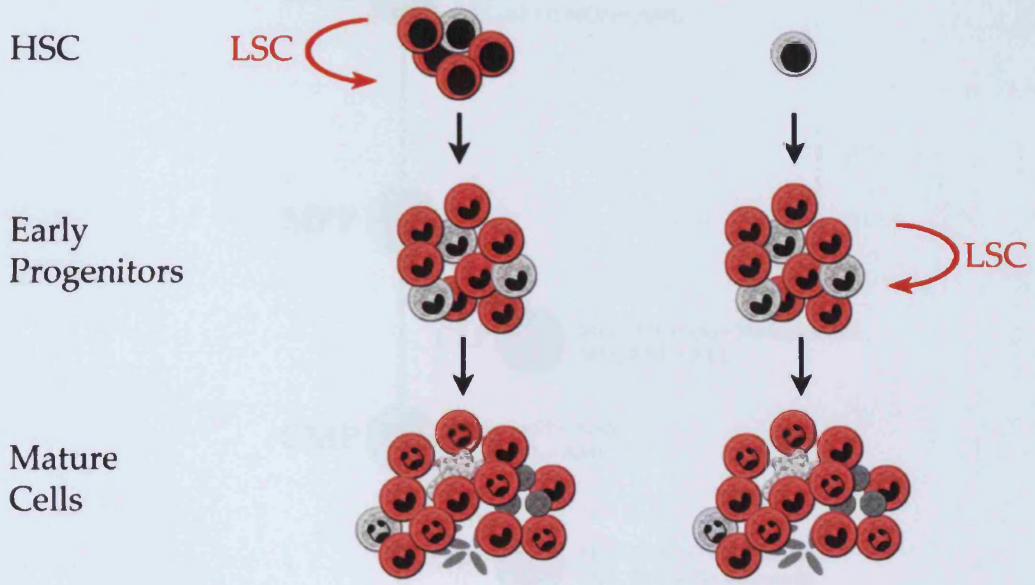


Figure 1.6: Site of origin of the leukaemic stem cell

The HSC is an attractive cell for the transformation into LSC, already having the ability to self renew and having longevity, allowing enough time for the accumulation of genetic events required for transformation. However the LSC could also arise from an early progenitor or mature cell that regains the ability to self-renew. Regardless of the site of origin, the bulk of the blast cells resulting would be the same. (Adapted from [3])

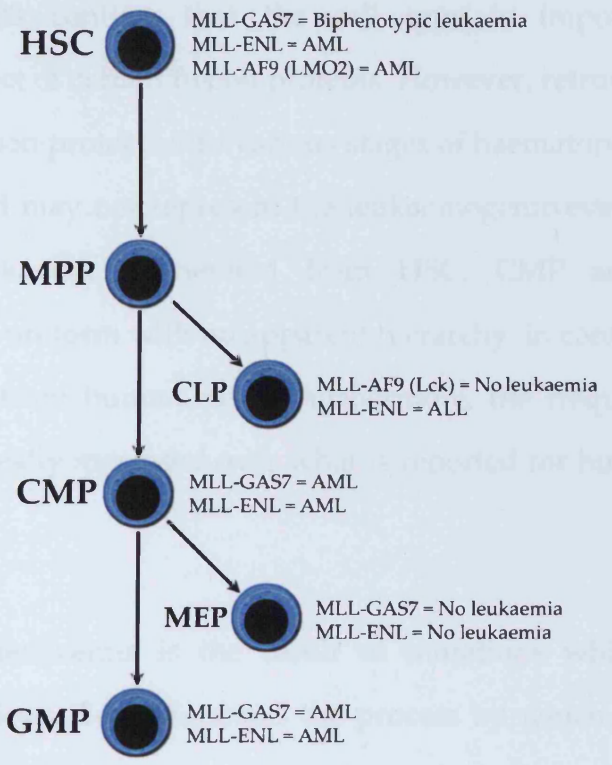


Figure 1.7: Relationship between cell type and leukaemic potential of various fusion proteins

Although some fusion proteins can indeed transform most stages of haematopoietic cell development, certain fusion proteins are only capable of transforming particular cell types or produce different leukaemias in different cell types. *MLL-ENL* can induce an AML in HSCs, CMP and GMP, but not MEPs. *MLL-AF9* can cause AML in HSCs, this fusion protein cannot transform CLPs. *MLL-GAS7* can cause only AML in CMP, GMP and MEP populations, but causes bi-phenotypic leukaemia when HSCs are transduced. These experiments demonstrate that the cell type is very important for leukaemic transformation. (Adapted from [46])

These results confirm that the cell type is important for the leukaemogenic effect of certain fusion proteins. However, retroviral-mediated transduction of fusion proteins into various stages of haematopoiesis is a very artificial model and may not represent the leukaemogenic event *in vivo*. The phenotype of leukaemias generated from HSC, CMP and GMP cell populations is very uniform with no apparent hierarchy, in contrast to what is observed in AML from human cases. Furthermore, the frequency of AML initiating cells is greatly increased over what is reported for human AML cell populations[47] [48].

Ultimately, leukaemia is the result of mutations which deregulate normal haematopoiesis. To understand the process by which mutations are able to achieve this, the types and patterns of mutations that are involved must be examined, along with studying when and how mutations arise and their functional impact on normal haematopoiesis.

LEUKAEMOGENESIS – TYPES AND PATTERNS OF MUTATIONS

CANCER IS CAUSED BY AN ACCUMULATION OF GENETIC EVENTS OVER TIME

Cancer arises from an accumulation of genetic and epigenetic changes in oncogenes and tumour suppressor genes that free cells from the mechanisms that control normal cell proliferation. The idea that cancer is caused by a series of successive somatic mutations has been supported by study of cancers such as retinoblastoma, in which the age of incidence is consistent with the requirement of two mutations or "two hits" for cancer development[49] [50]. The fact that AML is a disease that primarily affects the elderly supports the idea of damage accumulating over time. In AML this damage occurs at the very least in the form of chromosomal aberrations as

well as gene specific mutations.

Chromosomal aberrations

Unlike solid tumours which tend to have complex chromosomal abnormalities due to chromosome instability, the changes that are observed in haematological malignancies are often single translocations or numerical aberrations.

Gene and chromosomal deletions may involve a small part of the chromosome, such as the short or long arm (eg 5q-) or the loss of an entire chromosome (monosomy 7). Losses most commonly affect chromosomes 5, 6, 7, 11, 20 and Y. Gains are common in chromosomes 8, 12, 19, 21 and Y. Gene amplifications are not a common feature of haematological malignancies.

While numerical changes in AML are not unusual, balanced translocations of chromosomes, where no material is gained or lost, are the most common chromosomal aberrations observed. To date, there have been approximately 300 balanced translocations described in AML[51]. There are two main mechanisms by which balanced translocations may contribute to malignant change, either by overexpression of a normal cellular gene due to the fusion of the promoter region from one gene to the protein coding region of another, or by the fusion of two genes to generate a chimeric gene that encodes a novel 'fusion protein'. In AML, balanced translocations exclusively result in fusion proteins.

Fusion proteins can act by a variety of mechanisms in AML. Resultant fusion proteins may affect transcriptional control such as in chronic myeloid leukaemia (CML) and acute promyelocytic leukaemia (APL). The first

balanced chromosomal rearrangement of t(9;22), dubbed the Philadelphia chromosome, was observed in CML and causes a fusion of the genes *BCR* and *ABL* resulting in constitutively elevated tyrosine phosphokinase activity and subsequent deregulation of cell cycle control[52 53].

APL (AML FAB type M3) is characterised by the t(15;17) translocation where the promyelocytic leukaemia gene *PML* on chromosome 15 is fused to the retinoic acid receptor α gene, *RARα* on chromosome 17. The resultant *PML-RARα* fusion protein functions as a transcriptional repressor whereas the wild-type *RARα* is an activator[54]. However, not all translocations have dedicated partners; some genes such as the transcription factor *MLL* have many different fusions partners[55]. The functional consequence of the majority of fusion proteins is a block in differentiation. In the case of *MLL* and *RARα*, the primary function of various fusion partners is to promote dimerisation and/or homo-oligomerization of the transcription factors[56 57].

Balanced translocations are often found in leukaemia, however it is uncertain whether these translocations alone are sufficient to cause overt leukaemia. Low levels of the *BCR-ABL* transcript are detectable in normal individuals[58] indicating that while the translocation is important, it is not the only factor responsible for causing CML. Furthermore, a study on identical twins that developed childhood ALL demonstrated that the twins harboured the same *TEL-AML1* translocation *in utero*, but had variable latency periods before the onset of overt disease, indicating once again that additional genetic events are needed before the development of overt leukaemia[59].

Normal karyotype (NK) AML

While chromosomal aberrations, especially balanced translocations, are

well characterised in AML, they are often unable to solely account for the formation of the disease. The 10 most common balanced translocations occur in approximately 15% of AMLs. The other rare recurrent balanced translocations are observed in only 5% of cases. Unique structural changes, numerical changes in chromosome number and complex abnormalities each contribute to approximately 10% of AMLs respectively. This leaves approximately 50% of AMLs with no detectable cytogenetic abnormalities (**Figure 1.8**)[51].

Normal karyotype patients are generally classified as intermediate risk because as a group their CR rates, relapse risk and survival are worse than those of treated patients with favourable cytogenetics, but better than those with unfavourable cytogenetics[10] [13]. However, there are patients within the intermediate group that have poor or favourable prognosis. The lack of overt chromosomal lesions makes it difficult to stratify normal karyotype patients into specific prognostic groups and to monitor MRD.

Gene specific mutations

Another form of genetic damage that is observed in leukaemia is gene specific mutations. The completion of the human genome sequence has made every gene accessible to mutation analysis. Gene specific mutations can alter both normal proliferation and differentiation. Some result in constitutive activation of tyrosine kinases and their subsequent pathways, or alter the function of transcription factors. Genes commonly mutated in AML include *FLT3, NPM, CEBPA, RUNXI (AML1), MLL, cKIT* and *RAS.*

The FMS-like tyrosine kinase 3, *FLT3,* encodes a membrane-bound receptor tyrosine kinase which is important in haematopoietic stem

/progenitor cell survival and proliferation by signalling through the RAS pathway. It is mutated in about one third of AML patients and is also noted to be mutated in ALL and MDS, making it one of the most frequently mutated genes in haematological malignancies. *FLT3* mutations present either as internal tandem duplications (ITD) of the juxtamembrane domain (30%) or by point mutations usually involving the tyrosine kinase domain (TKD) (7%). Both types of mutation constitutively activate FLT3[60] [61]. Due to its constitutive activation in the mutated form, inhibitors of FLT3 are being developed as therapy for use in AMLs with *FLT3* mutations[62].

Heterozygous mutations in the nucleophosmin gene *(NPM1)* have recently been described as the most frequent genetic abnormality in AML, occurring in approximately 50-60% of normal karyotype adult AMLs[63]. Transgenic mice heterozygous for *Npm1* develop a haematological syndrome with features of a human myelodysplastic syndrome[64] indicating that *NPM1* gene mutations are involved in leukaemogenesis although the underlying mechanisms are not yet fully understood.

CCAAT enhancer binding protein alpha, *CEBPA,* is a transcription factor that is involved in myeloid differentiation and proliferation control. Mutations in this gene have been implicated in familial cases of AML; and also occur in approximately 10% of somatic cases[65]. The function and mutations of this gene will be described in greater detail subsequently.

GENE SPECIFIC MUTATIONS IN PROGNOSIS OF AML

The genetic aberrations causing AML can also provide valuable information in regards to prognosis and treatment of the disease. AMLs with translocations involving the core-binding factors (typically t(8;21) and inv(16))

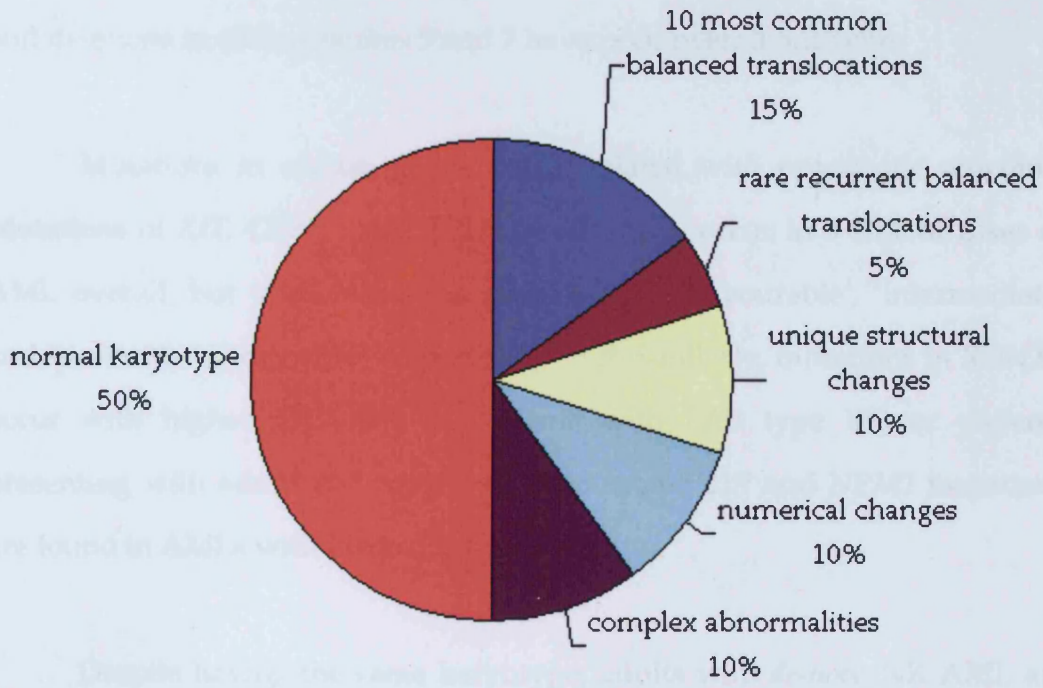


Figure 1.8: Karyotype breakdown of AML

Approximately 50% of AML have no cytogenetic abnormalities. The ten most common balanced translocations observed in AML are detectable in approximately 15% of cases. Rare recurrent balanced translocations, unique structural changes, numerical changes and complex abnormalities account for approximately 35% of AMLs. (Adapted from talk given by F. Mitelman at Charterhouse Square, 2007)

tend to do well overall with therapy whereas AMLs with complex karyotypes and deletions in chromosomes 5 and 7 have poor overall outcome.

Mutations in certain genes are correlated with prognostic outcome. Mutations of *KIT, CEBPA* and *TP53,* for example, occur in 5-10% of cases of AML overall, but predominate in patients with 'favourable', 'intermediate' and 'poor risk' cytogenetics respectively[66 67 68]. Similarly, mutations in *RUNX1* occur with higher incidence in patients with FAB type M0 or patients presenting with additional copies of chromosome 21[69] and *NPM1* mutations are found in AMLs with normal karyotype[63].

Despite having the same karyotype, adults with *de-novo* NK AML are diverse in respect to the acquired gene mutations and gene expression changes. These alterations affect clinical outcome and may assist in selection of proper treatment. As mentioned previously, currently NK AMLs are classified as intermediate outcome AMLs, but research has demonstrated that the presence or absence of certain mutations can also be useful in segregating AMLs into good and poor prognostic groups[5 70] (**Figure 1.9**).

FLT3-ITD mutation in the heterozygous state is a recognized poor prognostic risk factor for AML. Additionally, the loss of the wild-type *FLT3* allele occurs in ~10% of patients with normal karyotype AML and is associated with a worse outcome[71]. Recent studies have also shown that the size of the internal duplication of *FLT3* may affect prognosis, with ITDs of over 40bp in length having poorer outcome than those with shorter ITDs[72].

Mutations in *NPM1* are observed primarily within the normal karyotype subgroup, occurring in approximately 50-60% of all NK AMLs. Various studies have demonstrated that patients with *NPM1* mutation, in the

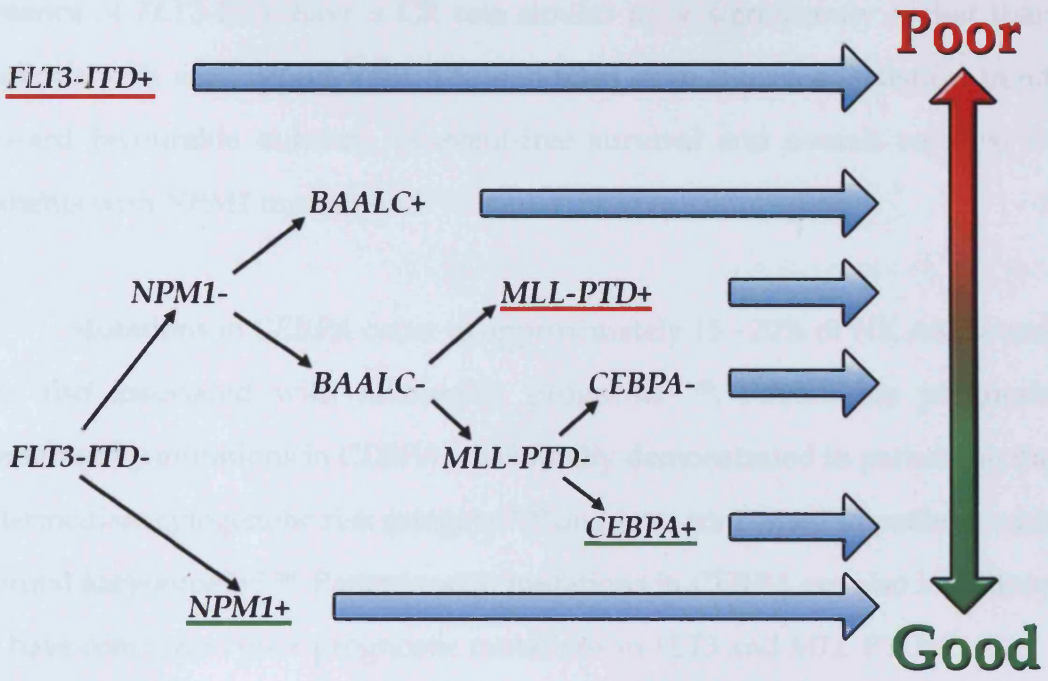


Figure 1.9: Mutations in normal karyotype AML can be useful in prognosis

Despite having the same normal karyotype, adults with *de-novo* NK AML are diverse in respect to the acquired gene mutations and gene expression changes. These alterations affect clinical outcome. The presence of *FLT3*-ITDs indicate poor prognosis. In the absence of *FLT3*-ITDs, the presence of *NPM* mutations and *CEBPA* mutations are indications of favourable prognosis, while the presence of *MLL*-PTDs and high *BAALC* expression are indicators of poor prognosis. (Adapted from [5])

absence of *FLT3*-ITD, have a CR rate similar to or significantly higher than patients with wild type *NPM1*. These studies have shown a statistical trend toward favourable outcome in event-free survival and overall survival in patients with *NPM1* mutations[73 74 75 76].

Mutations in *CEBPA* occur in approximately 15 - 20% of NK AMLs and are also associated with favourable prognosis[77 78]. Favourable prognosis conferred by mutations in *CEBPA* was initially demonstrated in patients in the intermediate cytogenetic risk category[79 80] and later confirmed in patients with normal karyotype[76 77 78]. Patients with mutations in *CEBPA* are also less likely to have concurrent poor prognostic mutations in *FLT3* and *MLL*-PTDs[77].

Partial tandem duplication in the *MLL* gene (*MLL*-PTD) was the first molecular alteration shown to impact on clinical outcome of NK AML. *MLL*-PTD is detected in approximately 5-10% of cytogenetically normal adults with *de novo* AML and is associated with *FLT3*-ITDs[74]. Patients with *MLL*-PTDs have significantly shorter complete remission durations than patients without the mutation[81 82].

The overexpression of specific genes in normal karyotype AML can also be used in prognosis, with overexpression of *BAALC*[78], *ERG*[83], and *WT1*[84] all conferring poor prognosis.

CLASS I AND CLASS II COOPERATING MUTATIONS

AML is a heterogeneous disease genetically, with over 300 different chromosomal translocations and mutational events described[85]. The complexity of the genetic data can present difficulties in both the clinic and in the understanding of the disease. Grouping mutations into broad catergories

based on their functional consequences is one way of simplifying the system.

Frequently genetically heterogeneous mutations target the same signal transduction and transcriptional pathways and therefore can be divided into groups. In an attempt to provide a more unified molecular theme to explain how different mutations can generate essentially similar phenotypes Gililland[86] proposed a "two-hit" model for leukaemogenesis. This is not to be confused with Knudson's two-hit model of cancer development as is applied to retinoblastoma where two genetic events are required to completely remove function of the Rb gene[50].

In Gililland's model, AML is the consequence of collaboration between at least two broad classes of mutation. Class I mutations confer proliferative and/or survival advantage to the cells, usually as a consequence of aberrant activation of signal transduction pathways. Class II mutations comprise mutations in transcription factors and transcriptional co-activators that are important for normal haematopoietic development with these mutations impairing differentiation (**Table 1.5**)[86].

Class I mutations are made up of a group of proliferative mutations that are well recognised in myeloid disease, including activated *RAS, FLT3, CKIT* or the *TEL-PDGFR* and *BCR-ABL* fusion genes. Although there are differences between individual tyrosine kinases, their functions result in similarities in activation of downstream effectors, such as activating STAT, PI3K and RAS signalling pathways. In murine models these genes cause myeloproliferation rather than true leukaemia[86].

Chromosomal translocations in AML often target transcription factors and transcriptional co-activators. The core binding factor (CBF) is often

Class I mutations	Class II mutations
BCR-ABL	CBFβ-MYH11
N-RAS	AML1-ETO
K-RAS	TEL-AML1
c-KIT (exon 8 and Asp 816)	PML-RARα
FLT3 (ITD and Asp 835)	NUP98-HOXA9
PTPN11	PU.1
NF1	C/EBPα
TEL-PDGFRβ	AML1
	AML1-AMP19

Table 1.5: Class I and Class II mutations in AML

Gilliland's "two hit hypothesis" postulates that the development of overt AML requires at least one mutation from each class. Class I mutations offer a survival or proliferative advantage to the cell, whereas Class II mutations are involved in blocking differentiation.

targeted, being involved in over a dozen translocations and point mutations. In many cases, these Class II mutations result in a block in differentiation pathways or apoptotic mechanisms, causing impaired haematopoietic differentiation. In contrast, many of these mutations in murine models cause expansion of the stem cell pool rather than overt leukaemia[86].

AML1-ETO is one of the most common fusion proteins observed in AML, however it is not able to induce leukaemia in experimental *in vivo* murine models on its own. Mice targeted to express *AML1-ETO* in the HSC compartment developed a non-lethal long-latency myeloproliferative syndrome but failed to develop acute leukaemia[87]. In a conditional *AML1-ETO* murine model the additional treatment with a mutagen was required before overt leukaemia developed[88]. To further support the two hit model, *AML1-ETO* was found to collaborate with *FLT3*-ITD in inducing acute leukaemia in a murine bone marrow transplantation model[89].

Gililland's two hit model suggests that a combination of both classes of mutations is required for leukaemogenesis, where both a block in differentiation and a proliferative and/or survival advantage are necessary to result in leukaemic transformation. Based on current large studies, the majority of AMLs have at least one genetic lesion described, however a proportion still remain with unknown genetic background.

LEUKAEMOGENESIS – LATENCY AND ORDER OF MUTATION IN DISEASE DEVELOPMENT

In most cases of leukaemia, multiple genetic events are required before the overt development of disease. Examination of the blast cells provides information about the pattern and types of mutations that make up AML,

however this does not give information about when the mutations arise and the order in which genetic events occur.

The simplest manner of examining order of mutation is by studying paired leukaemia samples. Changes in mutation status of genes between presentation and relapse may yield insight into the order by which they arose. A mutation which is maintained throughout the course of the disease is likely to have occurred early in the leukaemogenic process.

Progression of other haematological disorders to overt AML can help to demonstrate the requirement for accumulation of genetic damage for the development of disease. The progression from chronic phase to blast crisis in CML has been known to be associated with the gain of other mutations such as t(3;21)[90] [91]. *AML1-ETO* fusion followed by a subsequent *FLT3*-ITD has been observed with the progression of MDS to AML[92], which is also observed in murine models expressing these mutations[89]. Continuing with murine models, the co-expression of pairs of mutations such as *BCR-ABL* and *AML1-EVI1*[93], or *PML-RARA* and *FLT3* mutant[94] are necessary for the development of AML. Latency in *PML-RARA* knock-in mice before their development of APL suggests that the acquisition of new genetic lesions as necessary second events[95]. In humans, latency in AML development is best highlighted in the study of rare familial cases of the disease.

FAMILIAL AML

Familial cases of AML, while rare, can offer insight into the more common sporadic cases of the disease, especially in respect to the order by which mutation events occur. In addition to syndromic cases of familial AML which are associated with certain inherited genetic syndromes, such as

trisomy 8 and 21, single gene germline mutations can be responsible for predisposing families to AML[96].

Familial cases of syndromes such as neurofibromatosis and Noonan Syndrome have been useful for discovery of mutated genes that also yield the syndrome if they arise sporadically. Noonan syndrome is inherited in an autosomal dominant manner and maps to chromosome 12q24.1. Approximately half of patients with Noonan syndrome carry gain of function mutations in *PTPN11*, encoding the protein tyrosine phosphatase SHP-2 which is involved in embryonic development that modulate cell division and differentiation. Subsequently this gene has also been found to be mutated in AML[97].

Genes found to be involved in familial cases of AML include *RUNX1 (AML1)*, *CEBPA* and a yet uncharacterised gene located at 16q21-23.2[98]. Germ-line heterozygous mutations in *RUNX1* have been identified in familial AML[99]. Several families have also been described involving mutations in the transcription factor *CEBPA*[100 101 102]. These families exhibit long latency periods between birth and onset of the disease with up to 39 years before the development of AML[102]. All families described had normal karyotype. There are two main types of mutations that are observed in *CEBPA*, which will be discussed in greater detail in the *CEBPA* section of this introduction. Mutations in the 5′ end of the gene result in the translation of a truncated isoform of the protein, mutations in the 3′ end of the gene disrupt DNA-binding and dimerization domains in the C-terminus of the protein[65]. In familial cases of AML involving *CEBPA*, the germline mutation results in the truncated protein while additional C-terminal mutations are acquired somatically later in life. This manner of elucidating the order of mutations from familial cases can aid in understanding the molecular processes by

sporadic cases of AML develop.

LEUKAEMOGENESIS – FUNCTION OF MUTATIONS

Prospective mutation screening for genes involved in AML often arises due to their function, whether it is an important player in differentiation or regulation of proliferation. Therefore besides understanding the types of mutations present and the mechanisms whereby combinations of mutations result in the development of AML, understanding how these mutations function is equally important.

DISRUPTION OF DIFFERENTIATION – THE ROLE OF TRANSCRIPTION FACTORS IN AML

Mutations that disrupt transcription factors in AML are classified as Class II mutations which block mature differentiation of haematopoietic cells. AML research therefore focuses on the myeloid branch of development, where the transcription factors *PU.1, GATA1,* and *CEBPA* play key roles (**Figure 1.10**).

Concentration of transcription factors

Besides simple gain or loss of function in transcription factors, the concentration of factors drives normal haematopoiesis and leukaemogenesis. A small group of transcription factors with mostly lineage-restricted expression have been demonstrated to play a crucial role in the control of normal haematopoiesis[103]. Amongst the best-studied examples are *PU.1, CEBPA, AML1, GATA-1, c-myb* and *SCL/Tal-1.* These transcription factors all regulate a broad range of target genes, which drive differentiation along different developmental pathways[104]. Profound haematopoietic defects were observed when the genes encoding these factors are knocked out in mice[105, 106] .

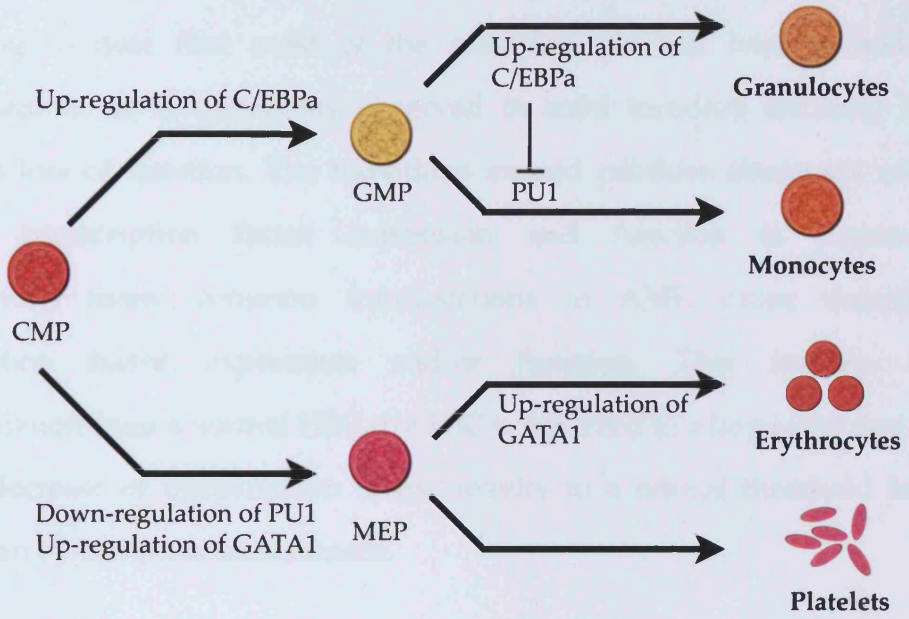


Figure 1.10: Transcription factor involvement in myeloid development

The transcription factors *PU.1*, *GATA1*, and *CEBPA* are key players in the differentiation of the myeloid branch of the haematopoietic system. The common myeloid progenitor (CMP) differentiates into a granulocyte/macrophage progenitor (GMP) with the upregulation of *CEBPA* and is driven towards the megakaryocyte/erythroid progenitor (MEP) by the upregulation of *GATA1* and downregulation of *PU.1*. Terminal differentiation is also influenced by these transcription factors, with upregulation of *CEBPA* resulting in granulocytes and upregulation of *PU.1* in monocytes from the GMP. The upregulation of *GATA1* from the MEP results in the production of erthyrocytes. These transcription factors are involved in complex regulation pathways, which are beyond the scope of this thesis to discuss. (Adapted from [107])

Although transcription factors are frequently mutated in AML, it is interesting to note that most of the mutations do not lead to biallelic null-mutations, as is commonly observed in solid tumours resulting in a complete loss of function. The mutations instead produce situations where residual transcription factor expression and function is preserved. Furthermore, many common translocations in AML cause decreased transcription factor expression and/or function. This implies that transformation from a normal HSC to a LSC is achieved in a large part due to a graded decrease of transcription factor activity to a critical threshold level, rather than by complete abolishment.

CCAAT Enhancer Binding Protein alpha (*CEBPA*)

As has been previously mentioned the transcription factor *CEBPA* plays a key role in myeloid differentiation and is frequently found mutated in NK AML as well as underlying familial cases of AML. Due to the type of mutation progression observed within familial cases and evidence of similar mutations in somatic cases, we aimed to further study the function of mutations in this gene.

C/EBPα is the founding member of a family a group of basic region leucine zipper transcription factors that all bind to a similar DNA consensus motif. In mammals six identified C/EBPs (α, β, γ, δ, ε, ζ) are expressed in a tissue-restricted manner[106]. These proteins share related N-terminal transactivating domains, basic DNA-binding regions and C-terminal leucine-zipper protein interaction domains. C/EBP family members are able to homo and heterodimerise via the leucine-zipper domains.

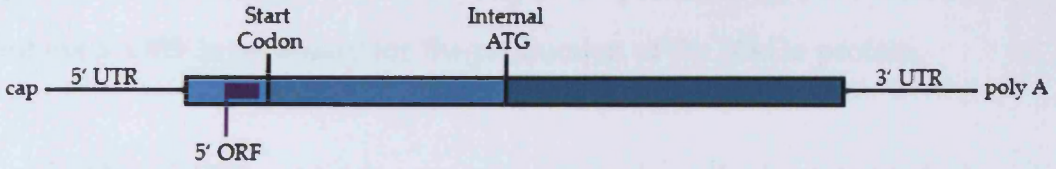
The protein C/EBPα is expressed in many cell types and plays crucial roles in hepatocyte and adipocyte development, with highest concentrations of the protein occurring in terminally differentiated cells. C/EBPα is also expressed in intestine, lung, adrenal gland, breast, ovary and placenta tissues[108]. In haematopoiesis, C/EBPα is expressed in myeloid cells and is a key regulator of granulocytic differentiation[65].

GENE AND PROTEIN STRUCTURE

CEBPA is a single exon gene located on the minus strand of chromosome 19q. Although the gene is intronless; tightly regulated alternative start codon usage results in the generation of two protein isoforms, a full length 42kDa isoform and a truncated 30kDa isoform. The 30kDa isoform contains the same carboxyl terminus as the full-length 42kDa form, but lacks the first 117 amino acids. *CEBPA* mRNA consists of a short 5′ untranslated region (UTR) containing a small 5′ open reading frame (ORF), the coding region and a large 3′ UTR (**Figure 1.11a**)[109, 110].

The translation of the different sized isoforms of C/EBPα is governed by ribosome scanning under the control of the 5′ ORF. This ORF is situated out of frame with the coding region of *CEBPA* and produces a small peptide (8 amino acids) that terminates just prior to the start codon for *CEBPA*. Ribosomes are loaded onto mRNA at its 5′ end and slide along until they reach an appropriate Kozak sequence, where they pause and translation begins. When the small 5′ ORF is translated, termination occurs just before the start codon and does not allow for sufficient time for repriming the ribosome once again. Thus the ribosome continues to travel along the gene until it reaches the second internal start codon and begins translation from there. Often the ribosome will skip the 5′ ORF due to a less optimal Kozak sequence, and

a)



b)

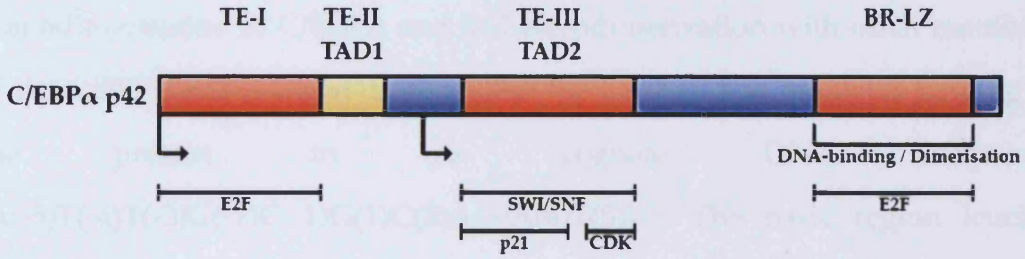


Figure 1.11: *CEBPA* gene and protein

a) The translation of the different sized isoforms of C/EBPα is governed by ribosome scanning under the control of the 5′ ORF. This ORF is situated out of frame with the coding region of *CEBPA* and produces a small peptide (8 amino acids) that terminates just prior to the start codon for *CEBPA*.

b) The basic region leucine zipper (BR-LZ) DNA-binding domain has been shown to mediate protein–protein interactions with other transcription factors involved in lineage-specific gene regulation (such as GATA1, PU.1, ETS1 and RUNX1) and growth control (such as c-JUN and E2F). N-terminal transactivation domains mediate interactions with the transcriptional machinery (TBP/TFIIB; CBP/p300; SWI/SNF). (Adapted from [65])

translate from the start codon, leading to the production of the 42kDa protein, but the 5′ ORF is necessary for the production of the 30kDa protein.

The C/EBPα protein contains several well characterised domains (**Figure 1.11b**). The C-terminal region of the protein consists of a basic region leucine zipper domain, in which the leucine zipper mediates homodimerisation of C/EBPα and its heterodimerization with other members of the C/EBP family. Upon dimerisation the basic region mediates binding of the protein to its cognate DNA sequence (A(-5)T(-4)T(-3)G(-2)C(-1)G(1)C(2)A(3)A(4)T(5))[110]. The basic region leucine zipper domain has also been demonstrated to be involved in protein interactions with other transcription factors involved in lineage-specific gene regulation (including E2F, GATA1, ETS1, PU1, c-JUN and RUNX1[65]). Distinct N-terminal transactivation domains (defined as TE-I, TE-II and TE-III by Nerlov and Ziff[111]; TADI and TADII by Friedman and McKnight[112]) mediate interactions with transcriptional machinery. The TE-I/TE-II or TADI domain, which is missing in the 30kDa isoform has been shown to be involved in E2F repression and interaction with transcription machinery TBP/TFIIB and CBP/p300. The TE-III or TADII domain which is present in both isoforms has been found to interact with p21, the chromatin remodelling proteins SWI/SWF, cyclin dependant kinases CDK2 and CDK4. Several pathways have been implicated as the means by which C/EBPα mediates cell cycle arrest and proliferation, including p21, cyclin-dependent kinases and the E2F complex via c-Myc. The 30kDa isoform of C/EBPα lacks the domains required to mediate cell growth[113].

CEBPA IN LEUKAEMOGENESIS

Mutations in *CEBPA*

Mutations in *CEBPA* have been implicated in AML, most often in association with FAB types M1 and M2, although it has also been found in M4 and M5 types. Mutations occur in approximately 10% of all AMLs and are associated with normal karyotype AML. Mutations in *CEBPA* tend to confer favourable prognosis[114], while low levels of its RNA expression have been noted in AML where it may reflect adverse prognosis. Although mutated *CEBPA* is primarily observed in AML, in rare cases it has been found to be mutated in MDS, lung and prostate tumours. Down-regulation of *CEBPA* has also been observed in blast crisis chronic myeloid leukaemia, lung, breast and liver cancers[115] [116] [117].

Mutations in *CEBPA* tend to cluster to two regions. The first group affects the N-terminal region of C/EBPα. These mutations are often insertions or deletions which cause frameshifts which cause premature truncation of the protein. In this case, translation is reinitiated at an internal ATG and the 30kDa protein. Secondly, in-frame or missense mutations occur within the C-terminal region of C/EBPα, disrupting the basic zipper region and thus affecting DNA binding, protein interactions, as well as homo and heterodimerization with other C/EBP family members[113].

Multiple mutations in *CEBPA* are common and often biallelic, although the allelic frequency of the mutations can change over the course of the disease[118] [80] [119]. Furthermore, the occurrence of both mutations on the same allele has been observed. Mutations are generally maintained between presentation and relapse and therefore may be useful for monitoring MRD.

Regulation of *CEBPA* expression in AML

In AML *CEBPA* is rarely involved in translocations with other genes. However, in B-Cell Precursor Acute Lymphoblastic Leukaemia (BCP-ALL), along with other members of the C/EBP family of proteins, *CEBPA* has been found to be translocated to the immunoglobulin heavy chain locus t(14;19)(q32;q13)[120]. The translocation does not affect the coding region of *CEBPA* but results in the downregulation of *CEBPA* expression. Additionally the gene retains its germline sequence.

Regulation of *CEBPA* expression is also commonly observed in AML and is ultimately heavily influenced by other common translocations. *CEBPA* expression is downregulated in the presence of fusion protein *AML1-ETO* indirectly by inhibiting positive autoregulation of the *CEBPA* promoter[121]. Translation of the C/EBPα protein is suppressed by the fusion gene *AML1-MDS1-EVI1* via the activation of calreticulin[122]. Pericentric inversion of chromosome 16, inv(16)(p13q22), which fuses the *CBFB* and *MYH11* genes, with the latter encoding the smooth muscle myosin heavy chain (*SMMHC*) also suppresses translation of the C/EBPα protein via calreticulin[123]. The *BCR-ABL* fusion protein is able to suppress C/EBPα protein translation via inhibitory action of the poly(rC)-binding protein hnRNP E2[124]. Furthermore, not only do fusion proteins affect *CEBPA* expression, but mutations which constitutively activate *FLT3* also inhibit C/EBPα function via ERK1/2-mediated phosphorylation[125].

CEBPA expression levels were also found to be important in explaining the function of new genes implicated in leukaemogenesis. Retroviral transduction of Tribbles homolog 2 (*Trib2*) into primitive murine HSCs resulted in the onset of overt leukaemia in which C/EBPα was downregulated

in transformed cells[126, 127].

These examples demonstrate the importance of regulation of *CEBPA* activity in AML. Even if the gene itself is not mutated, it can still play a role in the leukaemogenic process.

MODELS OF *CEBPA* FUNCTION

Extensive study has been carried out on the function of C/EBPα using a variety of models, including mouse strains with genetic alterations in the *CEBPA* locus, transduced human cells and other animal models such as zebrafish.

Knock out murine strains with complete abolishment of C/EBPα expression results in perinatal lethality due to severe defect in glucose homeostasis[127]. The embryos suffer from differentiation defects in liver, lung and fat as well as lack of mature neutrophilic granulocytes[128 129]. To overcome the perinatal lethality, mouse strains with inducible knock out of C/EBPα results in an accumulation of blast cells in the bone marrow, blocked at the CMP to GMP transition, but without the development of overt leukaemia[130]. Within C/EBPα null murine HSCs, there is increased expression of the polycomb group gene *Bmi1* which enhances HSC self-renewal ability. These cells were able to demonstrate a competitive advantage over wild type HSCs[131].

Upregulation of C/EBPα in human cells with the potential to follow a granulocytic or monocytic differentiation pathway results in granulocytic differentiation, with a block on monocytic differentiation[132]. However, in the murine model, C/EBPα expression directs monocytic commitment of primary

myeloid progenitors[133].

The functions of the C/EBPα isoforms have also been studied although the data has not always concurred. Within a murine cell line, expression of the 30kDa isoform of the protein limits G-CSF receptor expression[134] which is necessary for myeloid development. The exclusive expression of the 30kDa isoform within the murine model results in increased proliferation of myeloid progenitor cells, resulting in either a myeloproliferative disorder or AML[115]. Furthermore, knock-in mice with a targeted mutation in the basic region that specifically inhibits the C/EBPα-E2F interaction, also developed AML symptoms[115]. This suggests that while disruption of *CEBPA's* cell cycle control is important for AML induction in mice, other *CEBPA* functions must be preserved.

The effect of the shortened isoform in the human system is not as well characterised. Expression of the 30kDa isoform in human CD34+ cord blood was shown to inhibit both myeloid and erythroid lineages[135]. In zebrafish, expression of the shortened isoform of C/EBPα results in an increase in primitive erythropoiesis[136].

From the numerous studies that have already taken place, it is clear that the transcription factor *CEBPA* plays a key role in leukaemogenesis. However the exact mechanisms by how it contributes to leukaemia still remain unclear. Furthermore *CEBPA* mutations underlying familial AML warrants further study of this gene and its mutations.

SUMMARY AND AIMS OF THESIS:

The importance of cytogenetics in AML is well established with

implications for patient prognosis, treatment and options for disease monitoring. It is also accepted that changes at the molecular level are equally significant. Mutation analysis of candidate genes has been performed in a validated panel of AML identifying novel gene mutations and showing a temporal relationship between gene mutation and the occurrence of acquired homozygosity without copy number change.

Quantitative genomic assays have been developed for five patients with *CEBPA* mutations and used to examine the hierarchy of mutation events in the leukaemic stem cell pool. The functional aspect of these *CEBPA* mutations were tested by transduction of lineage depleted cord blood cells with these mutations to examine how they affect normal haematopoietic stem cells and contribute to the development of AML.

CHAPTER 2: MATERIALS AND METHODS

PATIENT SAMPLES

STORAGE OF SAMPLES

Patient samples were collected and stored in the Centre for Medical Oncology, John Vane Science Centre, Charterhouse Square with informed consent. Samples arrived in either solutions of RPMI with heparin or in an EDTA solution to prevent clotting. Samples in EDTA were diluted with RPMI before processing. Whole mononuclears were enriched by density centrifugation. 5mL of lymphoprep (Axis-Shield PoC, Oslo, Norway) were added to a 15mL tubes (Corning, Sigma-Aldrich, Dorset, UK) and 10mL of the blood sample were layered on top with a plastic pipette. Samples were then centrifuged at room temperature for 25 minutes at 1500 rpm (Beckman GS6R, Beckman Coulter, Fullerton, USA) with no break. The layer of white blood cells was removed and diluted in 10mL of RPMI. The tubes were then spun at 1300rpm for 10 minutes, high brake. Supernatant was discarded and cells were gently resuspended in an appropriate amount of freeze mix (final concentration of 10% DMSO) and transferred to cryovials. Cryovials were stored in liquid nitrogen. Various checks were carried out to insure fidelity of the sample.

DNA ISOLATION – PHENOL/CHLOROFORM

Genomic DNA from patient samples was extracted by standard phenol/chloroform methodology. In brief, cells were thawed rapidly and diluted with ice cold PBS and then pelleted by centrifugation. The cells were then resuspended in 1-2 mL digestion buffer (recipe at end of this chapter)

with added proteinase K and incubated at 55°C for 12-18 hours.

After incubation, an equal volume of phenol was added to the cell solution, mixed thoroughly by shaking and then spun for 10 minutes at 3000 rpm. The aqueous supernatant was removed to a fresh tube without disturbing the interface. An equal volume of phenol/chloroform (ratio 1:1) was added, mixed well by shaking, and then spun for 10 minutes at 3000 rpm. Again, the aqueous supernatant was removed to a fresh tube, to which an equal volume of chloroform was added, mixed and spun. The supernatant was removed for a final time and DNA was precipitated by the addition of 3 M sodium acetate, pH 5.2, followed by 2 volumes of 100% ethanol. DNA was either immediately spooled out of solution or left to percipitate at -20°C for 30 min to overnight before being spun down. DNA was then washed with 70% ethanol and resuspended in TE buffer.

DNA ISOLATION – MC LYSIS BUFFER

In cases where there was insuffient number of cells for phenol/chloroform extraction genomic DNA was isolated via this method. Recipe for lysis buffer is provided at the end of this chapter. 50μL of buffer was added to cells. Cells were incubated at 60°C for one hour and then at 95°C for 5 minutes. DNA was stored at -20°C.

MUTATION SCREENING

SINGLE-STRAND CONFORMATIONAL POLYMORPHISM (SSCP) MUTATION SCREENING

Mutation analysis of *CCND3* was carried out by single-strand conformational polymorphism (SSCP) analysis (GenePhorII, Pharmacia

Biotech, Amersham, UK). Precast 24-well 0.43 acrylamide gels (ETC, Elektrophorese-Technick, Kirchentellinsfurt, Germany) were rehydrated with 20mL of SSCP Rehydration buffer, pH 8.3 (ETC, Germany), for an hour. Rehydration buffer was initially optimised by running a test sample using each of the three available buffers (pH 7.4, pH 8.3 and pH 9.0) and selecting the buffer system which yielded the clearest bands.

All patients samples with normal cytogenetics were screened for mutations in *CCND3*. Each exon of *CCND3* (1-5) was amplified by PCR (primers and conditions at described at the end of this chapter). 1.5μL of PCR product was mixed with 4.5μL of formamide denaturing buffer (recipe at end of chapter), heated at 95°C for 5 minutes and then placed on ice. 6μL of product was loaded into each well of the rehydrated gel. Electrophoresis was carried out at 12°C beginning with 80V for 20 minutes, and then 510V until completion. Gels were then stained using the PlusOne DNA silver stain kit (Pharmacia-Biotech, Amersham, UK) on a Hoefer automated gel stainer (Amersham Biosciences, UK) following manufacture's instructions. Samples which appeared different from the others were sequenced.

WAVE MUTATION SCREENING

PCR reactions were preformed as described at the end of the chapter.

Denaturing high performance liquid chromatography (DHPLC) analysis

DHPLC was carried out using a Transgenomic WAVE Nucleic Acid Fragment Analysis system and DNASep column (Transgenomic, Crewe, UK). The composition of buffer A was 0.1M triethylammonium acetate (TEAA); buffer B contained 0.1M TEAA and 25% acetonitrile. Analysis was carried out

at a flow rate of 0.9mL/min and a buffer B gradient increase of 2%/minute for 4 minutes. Start and end concentrations of buffer B were determined empirically for each fragment. The column was run at two temperatures for each fragment. Temperatures are listed at the end of the chapter.

PCR products were prepared for DHPLC by denaturing at 95°C for 5 minutes and then cooling slowly to 65°C to allow heteroduplex formation. Data analysis was by visual inspection of chromatograms, which were examined independently by two observers. Samples from which chromatograms were unusual were brought forward for sequencing.

Direct Sequencing

Direct sequencing was used to screen for mutations in *CEBPA, FLT3, PU.1, RUNX1, WT1* and *CCND3* in the aUPD study. The genes were amplified by PCR covered the coding regions indicated and directly sequenced.

DNA sequencing – Slab and Capillary Gel

Unincorporated primers and dNTPs were removed from PCR products by ultra-filtration using Montage PCR centrifugal filter devices (Millipore corp., Bedford, MA, USA). Using the PCR product as template, a dideoxy chain termination method was used[137], with fluorescently tagged terminators and AmpliTaq DNA Polymerase enzyme (PE Applied Biosystems, Foster City, CA, USA). PCR conditions are provided at the end of the chapter.

For sequencing with the ABI 377 sequencer (Applied Biosystems, Foster City, CA, USA) the sequencing gel was prepared as outlined at the end of the chapter (Slab Gel). The products of the sequencing PCRs were recovered by

ethanol/EDTA/sodium acetate precipitation. 2μL of 3M sodium acetate solution (pH 5.2), 2μL of 125mM EDTA and 50μL of 100% ethanol were added to each sequencing reaction, vortexed and left at -20°C for 15 minutes to precipitate the extension products. The extension products were then pelleted by centrifugation at 14000 rpm in a bench-top centrifuge (Eppendorf Centrifuge 5417R) for 20 minutes and the supernatant aspirated and discarded. The pellets were washed with 100μL of 70% ethanol and air-dried.

The pelleted sequencing reactions were re-suspended in 4μL MICROSTOP loading buffer (Microzone Ltd, Haywards Heath, UK), denatured at 95°C for two minutes and held on ice until loaded. 1.5μL of the denatured sequencing products were loaded per lane of a polyacrylamide sequencing gel, separated by electrophoresis and read by fluorescent detection on an ABI 377 sequencer (Applied Biosystems, Foster City, USA).

Sequencing reactions carried out on the Applied Biosystems Prism 3730 were purified using DyeEx 2.0 Spin Columns (Qiagen, Crawly, UK). Columns were vortexed to resuspend the matrix and then spun for 3 minutes at 750g. The sequencing PCR was then applied to the compacted column and again spun for 3 minutes at 750g. The purified samples were then dried using a vacuum centrifuge (DNA Speed Vac, DNA120; Savant). Samples were run on the Prism 3730 as a service provided by CRUK.

DNA SEQUENCE ANALYSIS

Sequencing data were analysed using the Editseq sequence editing, Seqman II sequence assembly/contig management and Megalign multiple/pair-wise sequence alignment software (DNASTAR Inc, Madison, Wisconsin, USA), or visually analysed by examining the sequencing traces in

the 4Peaks program (http://mekentosj.com/4peaks/).

DNA sequences generated were compared with the sequences deposited in the Genbank Database using the National Centre for Biotechnology Information (Bethesda, USA) BLAST database searching facilities via the internet (http://www.ncbi.nih.gov). The Ensembl database (http://www.ensembl.org) was also used for interrogation of the human genome.

REAL TIME PCR

The real-time PCR assays and their analysis were carried out using the ABI Prism 7700 Sequence Detector System (ABI/Perkin Elmer, Foster City, CA, USA). All the reactions were carried out in duplicate with the following conditions: an initial 50°C for 2 minutes and 95°C for 10 minutes, followed by 40 cycles of a two stage PCR, consisting of 95°C for 15 seconds and 60°C for one minute. β-2-microglobulin (β2M) was used as endogenous reference gene. Analysis was carried out using the software provided with the ABI Prism 7700.

All real-time PCR reactions were carried out using Taqman technology. A probe is conjugated with a fluorescent reporter dye (TAMRA) on its 5′ end and a quencher on the 3′ end. The probe anneals to a sequence between the primer set. As the DNA is amplified the polymerase degrades the probe and the fluorescent dye is released from the influence of the quencher and produces a signal.

CLONING

AGAROSE GEL VISUALIZATION

DNA fragments were separated for analytical and preparative purposes by agarose gel electrophoresis, using Flowgen or Horizon 58 (Gibco, BRL - Paisley, Scotland, UK) gel electrophoresis apparatus tank apparatus. The electrophoresis solution used was 1X Tris-Borate-EDTA (TBE) buffer with ethidium bromide at 0.5μg/mL. Agarose (Life Technologies, Paisley, Scotland, UK) diluted to 1% in 1X TBE buffer was melted and upon cooling ethidium bromide solution added to a final concentration of 0.5μg/mL. The gels were cast into trays and an appropriately sized comb was used to form the loading wells.

Samples were loaded into the wells with 1/10 volume of gel loading buffer. Gels were electrophoresed at 100V and the appropriate DNA size ladder (1kb ladder - Gibco BRL, Paisley, Scotland, UK) was run in parallel. DNA was visualised by transillumination with ultraviolet light and photographed as a permanent record.

RESTRICTION ENZYME DIGESTS

DNA was suspended at approximately 1μg/μL in the appropriate 1x buffer, and digested with 1-10 fold excess of the appropriate restriction endonuclease (New England Biolabs, Hitchen, UK). For non-end cutting, 1 unit of enzyme being sufficient to cut 1μg of DNA when digested for 1 hour, and 10 units for cutting near the ends of DNA under the same condition. The digests were incubated at the recommended temperature (usually 37°C) for 1-16 hours as appropriate, the endonuclease heat inactivated and the digestion products stored at -20°C. The products of digestion were analysed by

separation of an aliquot of the digest using agarose gel electrophoresis, as described above.

GEL PURIFICATION

Restriction enzyme digest products were run on a 1% agarose gel. The bands of the appropriate size were excised using a clean scalpel with all efforts to keep exposure to UV light minimal. Qiagen Gel Purification kit (Qiagen) was used to recover the digested DNA following manufacturer's protocol. Final elution volume was 30μL.

TA LIGATION

CEBPA was cloned into Topo 2.1 vector by TA ligation (Invitrogen). A final extension of 10 minutes at 72°C adds an A base overhang on the PCR products. The prepared vector contains T overhangs. The ligation reaction was carried out according to manufacture's instructions, with 0.5-4μL of PCR product ligated with 1μL of Topo 2.1 vector in saline conditions for 5 minutes at room temperature. 1μL of the subsequent reaction was transformed into Top10 cells (Invitrogen), as described below.

SHRIMP ALKALINE PHOPHATASE (SAP) TREATMENT

To prevent vector cut with a single restriction enzyme from reannealing during ligation reaction, the vector was treated with SAP enzyme (Promega, Madison, USA) to remove the phosphate group from the 5′ overhangs. 1unit/μg of DNA was incubated for a minimum of 15min at 37°C in 1x SAP reaction buffer. SAP was inactivated by heating to 65°C for 15min.

T4 READY-TO-GO LIGATION

Cut vector and insert was ligated using Ready-To-Go T4 DNA ligase (Amersham Bioscience, Little Chalfont, UK). 50ng of vector was annealed with an appropriate amount of insert to achieve a 2:1 molar ratio between vector and insert in a final volume of 20μL. The DNA mix was added to the ligase tube and incubated at room temperature for 3-5mins, mixed and incubated at 16°C for 30mins, followed by heat inactivation at 70°C for 10mins.

GATEWAY CROSSOVER

Large genes were transferred into the lentivirus vector (LNT/Sffv-cddB MCS, **Figure 2.1,** plasmid a kind gift from Dr Adrian Thrasher)[138] by the gateway cloning system (Invitrogen) which uses site-specific recombination properties of bacteriophage lambda. 150ng of destination vector (lentivirus vector) and 300ng of entry clone (pENTR1A vector) were incubated with the recommended amount of 5x LR Clonase Reaction Buffer (4μL) in a 20μL reaction along with 4μL of LR Clonase enzyme mix. Reactions were incubated at 25°C for 1 hour. 2μL of Proteinase K solution was added to terminate the reaction, incubating at 37°C for 10min. 1μL was transformed into Top10 cells.

CLONING STRATEGY FOR MUTANT FORMS OF *CEBPA* INTO LENTIVIRUS BACKBONE

A schematic of this cloning strategy is shown in **Figure 5.3.**

CEBPA from patient AML sample was cloned into a Topo 2.1 vector (Invitrogen) by first performing nested PCR on patient genomic DNA and using the TA cloning kit (Invitrogen). This approach was possible since *CEBPA* is an intronless gene. The plasmid was transformed into Top 10 cells, and

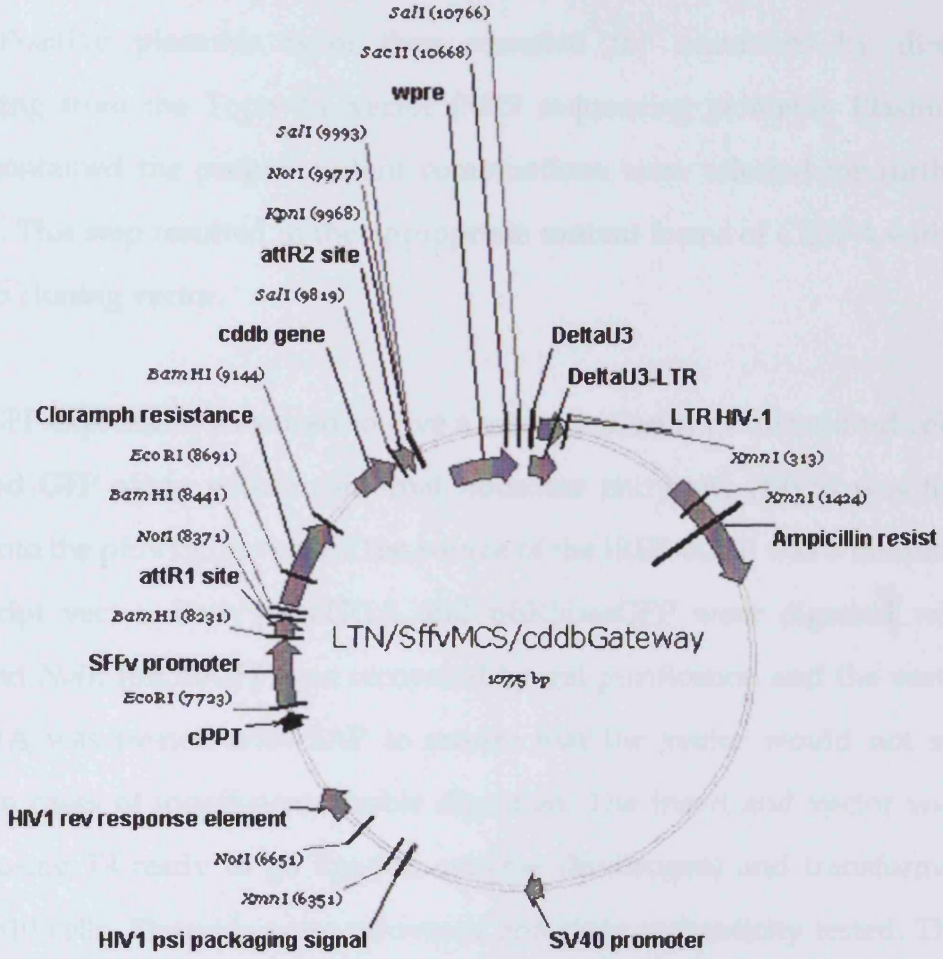


Figure 2.1: Lentivirus vector used for expression of mutant forms of *CEBPA*

Key features of this lentivirus vector include ampicillin resistance for perpetuating the plasmid in bacteria, a mammalian SFFV promoter, HIV elements to allow for recombination into the genome of the target cell and recombination sites (attR) for cloning of the target gene into the vector.[138]

colonies were picked based on blue/white selection. The plasmids were recovered by miniprep and tested for presence of the appropriately sized insert. Positive plasmids were then screened for mutations by direct sequencing from the Topo 2.1 vector (M13 sequencing primers). Plasmids which contained the proper mutant combinations were selected for further analysis. This step resulted in the appropriate mutant forms of *CEBPA* within the Topo cloning vector.

GFP expression was used to give a visual indicator of transduced cells. Enhanced GFP along with an internal ribosome entry site (IRES) was first cloned into the pENTR1A vector. The source of the IRES-eGFP was a modified pBluescript vector. Both pENTR1A and pBKSiresGFP were digested with *EcoRI* and *NotI*. IRES-eGFP was recovered by gel purification and the vector pENTR1A was treated with SAP to ensure that the vector would not self anneal in cases of insufficient double digestion. The insert and vector were ligated using T4-ready to go ligation enzyme (Invitrogen) and transformed into Top10 cells. Plasmids were recovered and their authenticity tested. This step produced an entry vector containing IRES-eGFP.

The mutant forms of *CEBPA* from the Topo plasmids were then cloned into the pENTR1A-ires-GFP plasmid. A FLAG tag was added to the C-terminal end of the protein to allow differentiation between endogenous C/EBPα and introduced forms of the protein. Restriction enzyme digestion sites were introduced to the ends of the gene to allow for directional cloning; (Forward primer 5′-agtctggatcctcgccatgccgggagaactctaac-3′ Reverse primer 5′-attatggatccttatttatcgtcatcgtctttgtagtccgcgcagttgcccatggcctt-3′) The gene was amplified from the Topo vector using a high fidelity Taq polymerase (TaKaRa). The pENTR1A-ires-GFP plasmid and *CEBPA* PCR product were digested with appropriate restriction enzymes, with the vector undergoing

additional treatment with SAP. Vector and insert were ligated and transformed into Top10 cells, the plasmids were recovered and tested for proper cloning. At this point, the entire length of *CEBPA* was sequenced to insure that no point mutations were introduced from the PCR steps. This step produced a plasmid with the mutant *CEBPA* gene along with GFP in the pENTR1A backbone.

The prepared pENTR1A-*CEBPA*-ires-GFP plasmid was recombined with the lentivirus vector, delivering the gene and GFP into the lentivirus backbone (**Figure 2.1**). Once again the entire length of the gene was sequenced to insure fidelity of the gene and mutation sequences. A mock construct was also made by recombining IRES-eGFP the lentivirus backbone. This final construct contained the mutant *CEBPA* gene along with IRES-eGFP within the lentivirus backbone.

BACTERIAL CULTURES

TRANSFORMATION INTO TOP10 CELLS

One to two microlitres of plasmid were transformed into 50 μL of chemically competent Top 10 (Invitrogen) cells according to manufacturer's protocol. Briefly, DNA was incubated with the cells on ice for half an hour and then heat shocked at 42°C for 30 seconds. Two hundred and fifty microlitres of SOC media was added to the cells and incubated with shaking at 37°C for 1 hour. Fifty and 200μL were plated onto LB agar plates containing the appropriate antibiotic and incubated overnight at 37°C. If the plasmids contained β-lactose reporter gene, X-gal (20mg/mL) was added to the plates for blue/white selection.

PLASMID PREPARATIONS

Bacterial cultures were grown in LB broth supplemented with either 50μg/mL of kanamycin or 100μg/mL ampicillin in a 37°C shaking incubator.

Miniprep

Plasmid recovery from small amounts of bacterial culture (3mL) was carried out using Wizard® Plus Minipreps DNA Purification System (Promega) following manufacturer's instructions. Final elution volume was 50μL of dd$H_2$0.

Maxiprep

For purification of up to 500 μg of transfection grade plasmid DNA, QIAGEN Plasmid Maxi Kit was used, following manufacturer's instructions.

Megaprep – endotoxin free

For purification of up to 2.5mg advanced transfection grade plasmid used in lentivirus production, QIAGEN EndoFree Plasmid Mega Kit was used, following manufacturer's instructions.

GLYCEROL STOCKS

Bacteria were stored at -80°C in 15% glycerol.

WESTERN BLOTTING

PROTEIN ISOLATION

Pelleted cells were resuspended in 100-200μL of cell lysis buffer (recipe

provided at end of chapter) and incubated on ice for 20 minutes. The solution was then spun at 4°C for 15 minutes at 14000 rpm and supernatant which contained the protein was collected. Protein concentration was measured by means of a chromatic assay in which a dye reagent (Biorad, Hercules, USA) changes colour in response to various concentrations of protein.

The standard curve was made by diluting commercial lyophilised bovine serum albumin (BSA, NEB) to concentrations ranging from 0-1mg/mL in double distilled water. 5μL of the standards and unknown samples were added to 250μL of 1:4 diluted Biorad dye reagent in a flat bottomed 96 well plate. Protein concentration was measured by plate reader at a wavelength of 595nm.

LOADING AND RUNNING THE GEL

For C/EBPα, 12% SDS-PAGE gel was chosen for good resolution of the 42 and 30 kDa proteins. 1mm gels were prepared (end of chapter for recipe) and assembled using Biorad mini-PROTEAN gel electrophoresis system (Biorad, Hercules USA). Protein was mixed with loading buffer (end of chapter for recipe), denatured by boiling for 5 minutes at 95°C and kept on ice until ready to load. Once loaded, the gel was run at 80V for 10 minutes and then at 120V until completion. Each time a molecular weight marker (broad range kalediscope prestained ladder, Biorad) was run along side samples to allow for the identification of bands.

TRANSFER AND BLOTTING

To continue the western blotting, the protein was transferred from the gel to a solid membrane. The separated protein was transferred onto a

Hybond-C membrane (Amersham Pharmacia) using mini Trans-Blot system (Biorad) in 1x transfer buffer, 4°C for 16h at 34V.

The membrane was blocked with 5% Marvel milk in TBST (1x TBS + 0.05% Tween, TBS recipe provided at end of chapter) for 1 hour at room temperature, on a rocking shaker. The membrane was then washed 3 times with fresh TBST, with 5 minutes for each wash. Primary antibody was added at optimized dilutions (1:1000 for all C/EBPα antibodies; Santa Cruz Biotechnology, Santa Cruz USA) to 5% BSA in TBST and applied to the membrane for 1 hour, room temperature on a rocking shaker, or for overnight at 4°C on a rotating shaker. The membrane was then again washed three times as described above. Secondary antibody (1:2500 dilution, Pierce Biotechnology, Rockford USA) conjugated to horse radish peroxidase was added to 5% milk in TBST at appropriate dilution and applied to the membrane for 1 hour, room temperature on a rotating incubator. The membrane was then finally washed 3 times with fresh TBST for 10 minutes each on a rocking incubator.

The westerns were visualised by enhanced chemiluminescent (Pierce Biotechnology) and exposed to film (Kodak, Switzerland) for various lengths of time. Film was developed using a Hoeffman automatic developer.

CELL CULTURE

MAINTANENCE OF ADHERANT CELLS

293T and HeLa cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM, Gibco Invitrogen) high glucose, supplemented with 10% fetal calf serum (FCS, Gibco Invitrogen) and 1% penicillin/streptomycin

(CRUK) in cell culture treated flasks (Falcon, BD Bioscience, Oxford UK) in a 5% CO_2 incubator. Cells were maintained at healthy growing densities. To split cells, media was removed from the cells by aspiration and cells were washed with PBS. Trypsin (CRUK) was added to the cells and incubated for 5 minutes at 37°C. FCS was added to stop the reaction and cells were spun at 1300rpm with high break applied. Cells were resuspended in media and counted on a haemocytometer. An appropriate number of cells were perpetuated in fresh media.

STEM CELL PURIFICATION

Stem cells were extracted from human umbilical cord blood. Informed consent was obtained from mothers awaiting caesarean section at the obstetric units of the Royal London Hospital, Whitechapel, London and the University College Hospital, Camden, London. Blood was extracted from the placenta and umbilical cords with a 20mL syringe and immediately supplemented with heparin (Sigma Aldrich, Dorset, UK) to a final concentration of 1mg/mL in 50mL falcon tubes. Tubes were transferred to the laboratory and stored overnight on a rocker at room temperature for extraction of stem cells on the next day.

The cord blood was diluted 1:4 with sterile PBS. Thirty-five mL of diluted cord blood was carefully layered on top of 15mL of Ficoll-paque plus, a solution of Ficoll 400 and Distrizoate Sodium with a density of 1.077g/mL (Stem Cell Technologies, Meylan, France) in a 50mL tube. Tubes were spun at 400g for 30 minutes with no brake applied. Mononuclear cells which form a layer at the interphase of serum and ficoll were collected with a Pasteur pipette into new tubes. Cells were spun and then resuspended in a suitable volume of PBS supplemented with 2% FCS (to be referred to as 2% FCS) and at least three

times the volume of ammonium chloride solution (Stem Cell Technologies, Meylan, France) for red cell lysis. The cells were stored on ice for five minutes, after which 2mL of foetal calf serum was added to stop cell lysis. Cells were spun again and resuspended in 2% FCS and counted on a haemocytometer.

LINEAGE DEPLETION

Mononuclear cells isolated from human cord blood were depleted of cells expressing mature lineage antigens using the StemSep column system from Stem Cell Technologies (Meylan, France). Cells were diluted to $8x10^7$ cells/mL in 2% FCS. 100μL of StemSep enrichment cocktail was added per mL of cells. The enrichment cocktail contains monoclonal antibodies for CD2, CD3, CD14, CD16, CD19, CD24, CD56, CD66b and Glycophorin A. An additional aliquot of 10μL/1mL cells of CD41 antibody was added to enhance depletion of platelets. After incubation on ice for 30 minutes 60μL/mL of thoroughly vortexed magnetic colloid was added. Following an additional incubation of 30 minutes on ice, cells were processed through a metal mesh column suspended in a strong magnetic field according to the manufacturer's recommendations. The flowthrough of the column, containing unmarked lineage negative cells was collected in 2% FSC. The cells were spun, counted and either frozen in FCS supplemented with 10% dimethyl-sulphoxide (DMSO, Sigma) or used immediately for transduction or transplantation. The column, containing the lineage positive cells, was washed with 2% FCS and the cells were either discarded, or frozen down and used as accessory cells in transplantation experiments.

TRANSIENT TRANSFECTION OF 293T CELLS

293T cells were transiently transfected with lentivirus plasmid

containing the *CEBPA* clones to test for protein expression. Poly(ethylenimine) (PEI 1mg/mL) was used as a transfection reagent, using a ratio of 1:1 between PEI and DNA. Transfection was performed in serum-free media (Optimem, Gibco Invitrogen). PEI was first mixed with 5% final transfection volume of media and left at room temperature for 5 minutes. DNA was added to 5% final volume of transfection. PEI and DNA were then mixed and left for 20 minutes at RT and then added to the cells. Cells were left in transfection media for 4 hours in normal incubation conditions. The transfection media was then aspirated and replaced with standard growth media. Media was changed 24h after transfection and cells were harvested either 48 or 72h post transfection by scrapping into ice cold PBS. For western blotting, cells were pelleted, supernatant removed and stored at -20°C until use. For other assays cells were removed from the plate by trypsin.

INTRACELLULAR STAINING

Cells were resuspended to a concentration of $1x10^6$ in 500µL of PBS. 500µL of RT 4% Paraformaldehyde was added and left to incubate for 15 minutes at RT (final concentration 2% PFA). Cells were washed by adding 3-4mL of PBS and spinning at 1300rpm for 7 minutes (low break). Supernatant was aspirated and cells were resuspended in 500µL of PBS. 500µL of 0.2% Triton-X100 (Tx) was added and cells were left to incubate at RT for 10 minutes (final concentration 0.1% Tx). Cells were washed twice as described above and resuspended in 50µL 2% FCS. Primary antibody (14AA anti-C/EBPα, Santa Cruz) was added to the cells to a final concentration of 1µg /million cells and incubated for 1 hour at RT. Cells were washed and resuspended in 50µL 2% FCS. Secondary antibody conjugated to a fluorochrome (antirabbit-APC) was added to a final concentration of 1µg/million cells and incubated at RT for 45 minutes in the dark. Cells were washed, resuspended in 500µL 2% FCS and

analysed by flow cytometry (FACSCalibur, BD Biosciences, FL1-H channel for GFP, FL4-H channel for APC).

LENTIVIRUS PRODUCTION

The viral vector used to transduce lineage depleted cord blood cells was produced by transient transfection of three plasmids into 293T cells, the self-inactivating transfer vector plasmid containing the gene of interest (**Figure 2.1**), a multideleted packaging plasmid and pMD.G which provides the viral envelope[139].

A total of 4 x 10^6 293T cells were seeded in as 175cm^2 flask 3 days prior to transfection. Cells were cultured in DMEM with 10% FCS, 1% penicillin and streptomycin in a 5% CO_2 incubator. 17.5μg of the envelope plasmid, 32.5μg of packaging plasmid and 50μg of transfer vector plasmid were precomplexed with polyethylenimine (PEI, Sigma) in a 1:1 ratio of DNA to PEI in 10mL of Optimem media at room temperature for 20 minutes. The DNA plus PEI complexes were added to the cells with the addition of 15mL of Optimem, resulting in a final volume of 25mL. In general 6 flasks were transfected with the same viral vector at one time. The cells were incubated at 37°C for 4 hours, after which the media was replaced with fresh supplemented DMEM. After 24 hours, the media was replaced with fresh media. After 48 hours the medium was harvested and cleared by filtering through 0.45nm filter. Virus was concentrated by spinning for 3 hours at 25,000rpm in a Beckman ultracentrifuge. After spinning, the supernatant was removed and 50μL Stemspan media (Stem Cell Technologies) added to the pellet, which was left to resuspend on ice for 1 hour. The virus was then aliquoted into 100μL fractions, with 30μL being saved for titration purposes. This harvesting and concentrating procedure was also followed for 72 and 96 hours following

transfection. If greater production was required, 12 flasks were used instead of 6.

To titre the virus, 50,000 HeLa cells were plated into each well of a 12 well tissue culture treated plate (Falcon) in a volume of 1mL normal growth media. Serial dilutions of the virus ranging from 1/100 to 1/100,000 for each day were prepared and added to the cells. After 72 hours the cells were analysed for GFP expression by flow cytometry on a FACScalibur (BD Bioscience, San Jose USA). Virus titre was calculated based on percentage of GFP positive HeLa cells.

TRANSDUCTION OF LIN- CELLS

Lineage depleted cord blood mononuclears (Lin- MNCs) were aliquoted into 96 or 48 well flat bottom tissue culture plates at a concentration of 5 x 10^5/mL in Iscove's Modified Dulbecco's Medium (IMDM, Gibco) supplemented with cytokines (SCF, IL-6, FLT3) and left to activate under normal growth conditions (37°C, 5% CO_2) for 4 hours.

Thirty virus particles were added for every Lin- MNC. If the change in the volume of the well was great, additional cytokines were added to maintain correct concentration. Cells were left to transduce with the virus for 16h under normal growth conditions. Following transduction, cells were checked for blebing (a sign of apoptosis) and washed twice in 2% FCS. Cells were counted again and used immediately for *in vivo* (NOD/SCID xenotransplantation) and *in vitro* (methylcellulose assays) experiments. Any remaining cells were analysed by flow cytometry (FACScalibur) for GFP expression to determine the efficiency of transduction.

METHYLCELLULOSE ASSAY

One thousand virus transduced Lin- MNCs were added to each mL of H4434 Methocult media (Stemcell Technologies) supplemented with 1% penicillin/streptomycin. Cells and media were vortexed gently to ensure even distribution of cells throughout the viscous media. 1mL of methylcellulose and cells were dispersed into a 35mm cell culture plate (Falcon) using a 1mL syringe and blunt end needle. All samples were plated in duplicate and maintained in high humidity conditions. Plates were left for 14 days under normal growth conditions and then colonies were scored by visualisation under an inverted microscope.

SCORING METHYLCELLULOSE ASSAYS

Methylcellulose assays only allow us to infer the cell of origin and are therefore retrospective in nature. By observing the type of colony present, the cell of origin can be determined. More primitive cells are able to give rise to larger colonies of mixed cell types. Colonies are categorised based on the morphology of the cells. **Figure 2.2** shows examples of the various colony types that have been observed in the assay. They can be assembled into the following 5 categories, BFU-E, CFU-E, GM, G and M.

BFU-E: Burst forming unit-erythroid

BFU-E colonies are composed of at least three clusters of tightly packed cells. The colonies vary in size and arise from primitive erythroid progenitors that have a high proliferative capacity. The cells remain nucleated and are distinguishable by a red colour produces as they become haemoglobinised (**Figure 2.2a, Figure 2.3**).

CFU-E: Colony forming unit-erythroid

CFU-E colonies consist of one or two clusters of erythroid cells and represent more mature erythroid progenitors than BFU-E. These have less proliferative capacity than BFU-Es (**Figure 2.2b**) and they can also be distinguished from myeloid cells based on the presence of red-colouration.

G: Granulocyte

Clonogenic progenitors of granulocytes give rise to a homogenous population of eosinophils, basophils or neutrophils. The colonies are colourless with individual cells smaller in size than macrophages (**Figure 2.2c**).

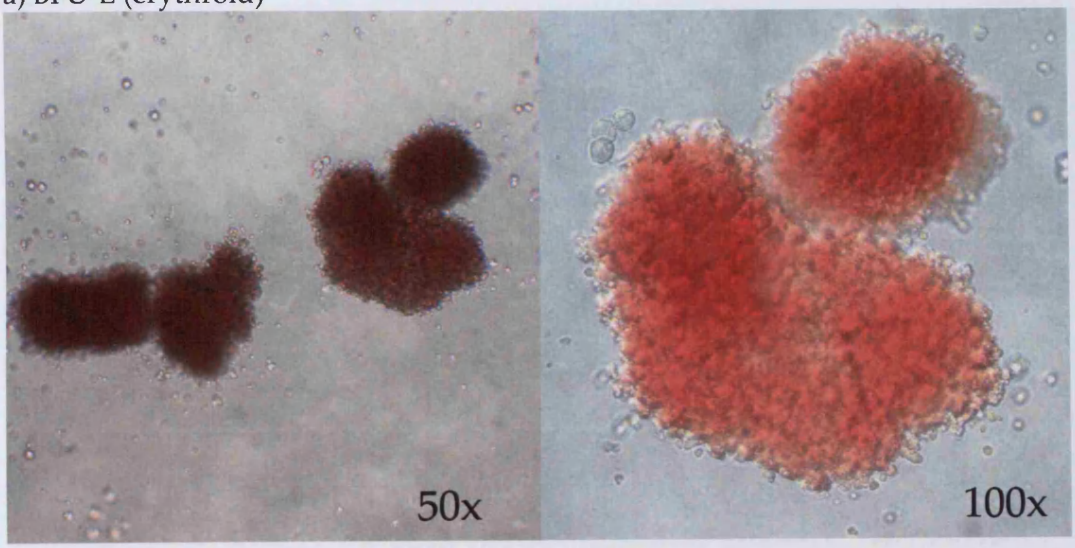
M: Macrophage

Homogenous, colourless macrophage colonies are composed of cells which are notably larger than granulocytes or erythrocytes. Macrophages also tend to grow in sheets rather than clusters distinguishing them from granulocyte colonies (**Figure 2.2d**).

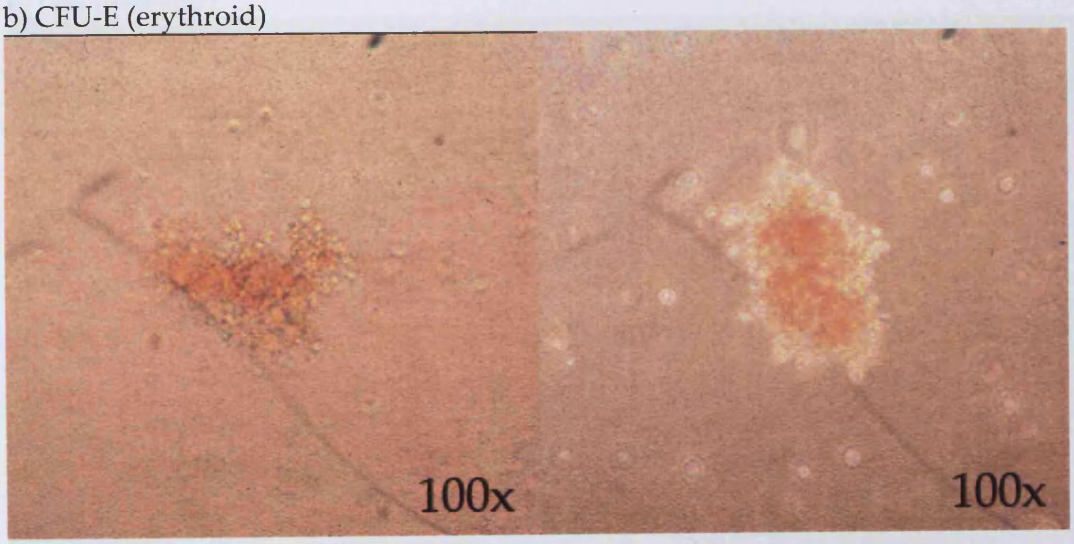
GM: Granulocyte, Macrophage

The progenitors of these colonies are able to give rise to a mixed population of macrophages and granulocytes and are thus of a more primitive cell. The colonies are colourless and can vary greatly in size. They are comprised of a tight centre of granulocytes surrounded by a halo of macrophages, although diffuse colonies are also observed (**Figure 2.2e**).

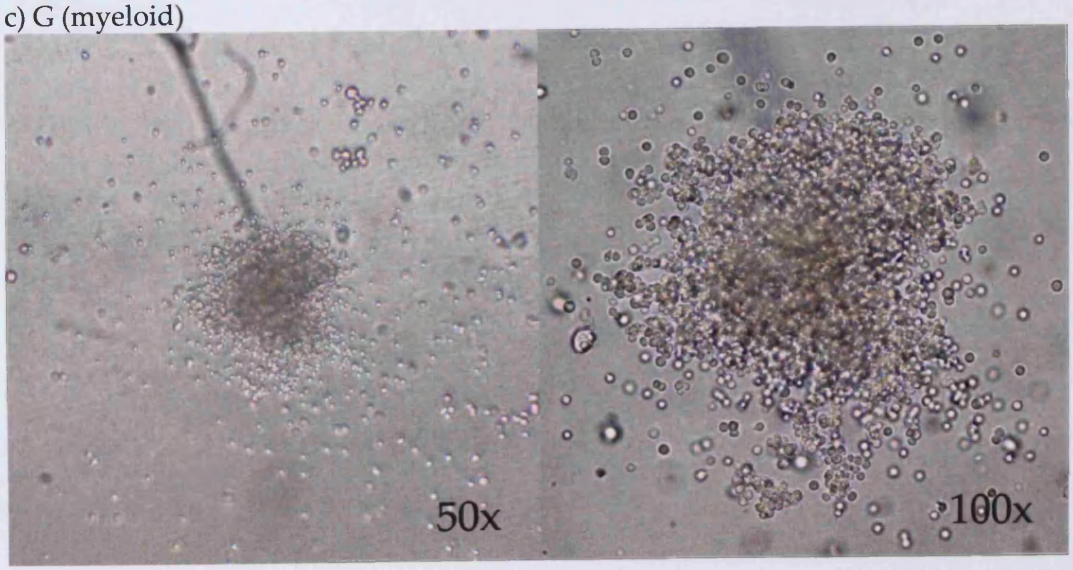
a) BFU-E (erythroid)



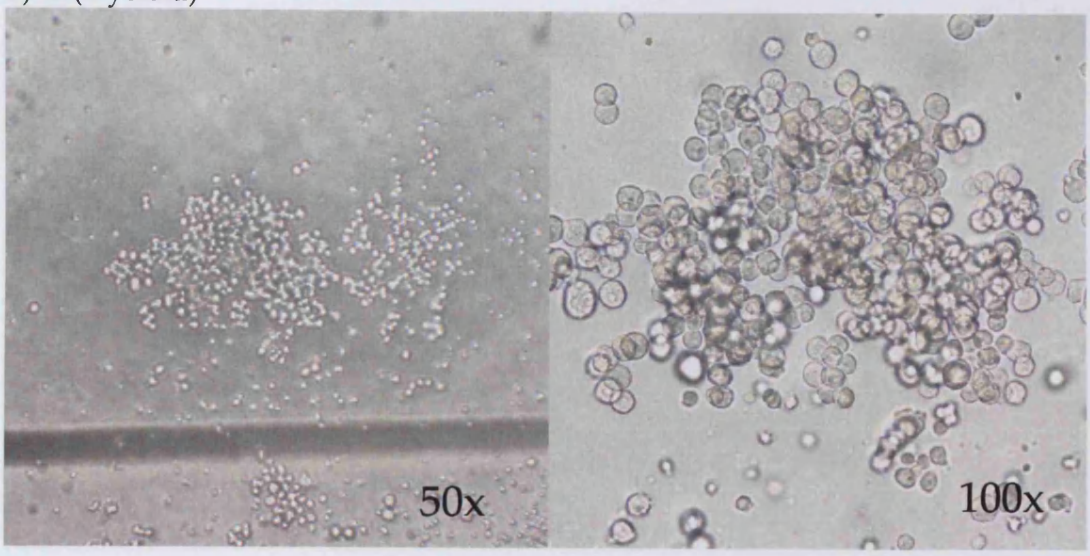
b) CFU-E (erythroid)



c) G (myeloid)



d) M (myeloid)



e) GM (myeloid)

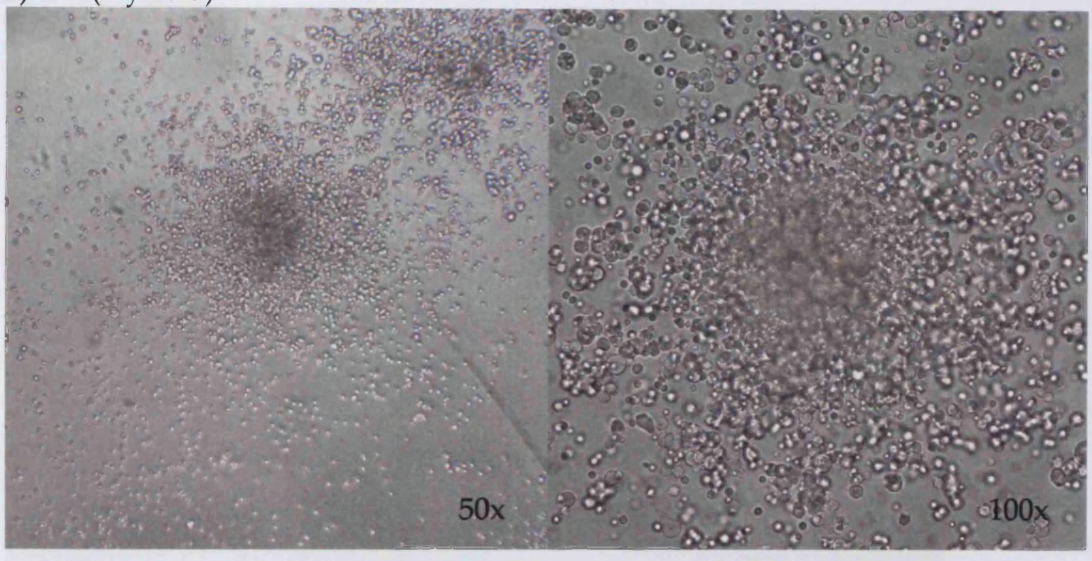
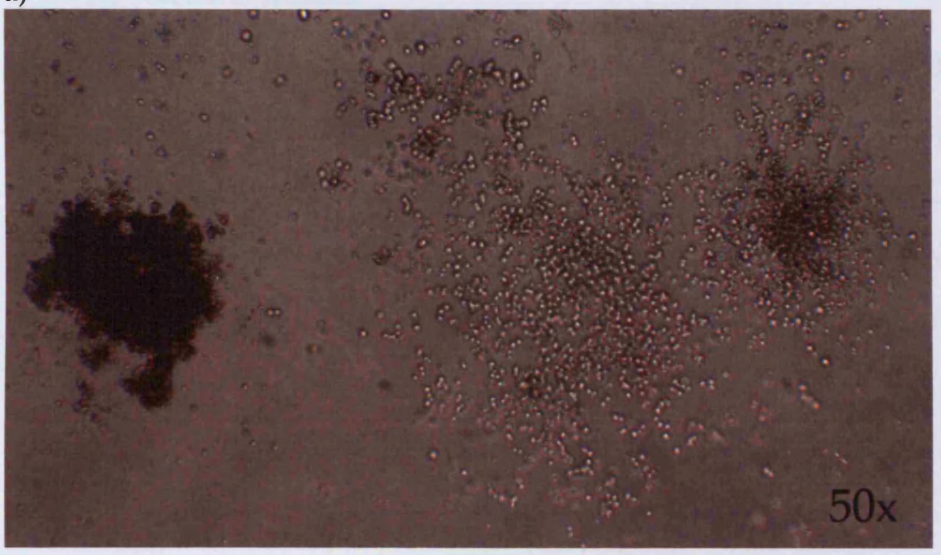


Figure 2.2: Examples of colonies types observed on methylcellulose plates

Colonies observed on methylcellulose plates are scored based on morphology. Each colony type observed in the assay is shown at a magnification of 50x and 100x. Please see text for descriptions of colonies.

a)



b)

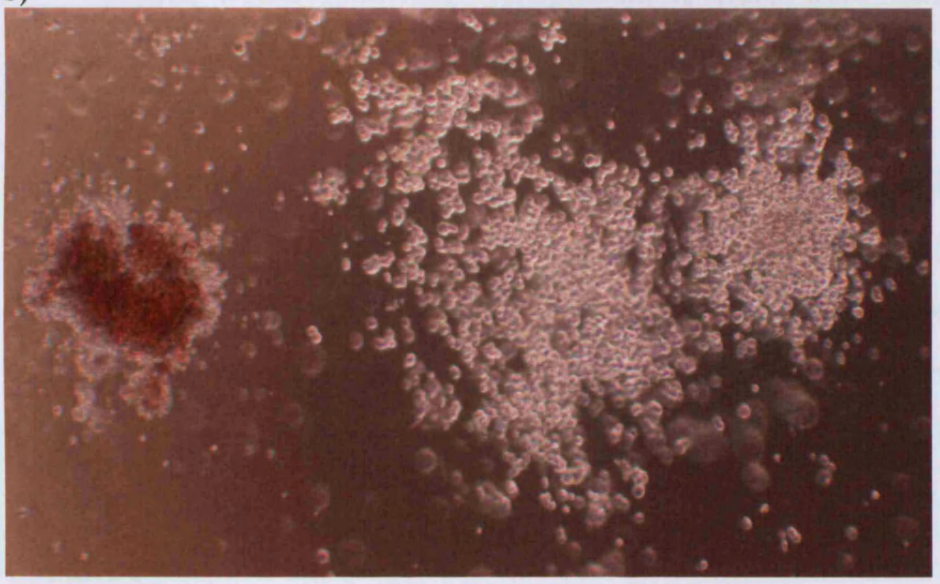


Figure 2.3: Scoring myeloid versus erythroid colonies

The simplest way to distinguish myeloid from erythroid colonies is through haemoglobin contained within erythroid colonies. Erythroblasts reach maturity by 10 to 12 days and can be distinguished by the reddish color displayed. After day 14, the colony may appear brownish due to the lysed erythroblasts.

a) Erythroid (left) and myeloid (right) colonies under normal light (50x)

b) The same colonies under a cone aperture of light to highlight the difference in colour between the colonies.

XENOTRANSPLANT MODEL

NOD/SCID and NOD/SCID β2 mice are maintained under sterile conditions in the animal facility at CRUK. Prior to engraftment experiments mice were sublethally irradiated (375 rads, cesium-137 source). Cells were injected via tail vein in a volume no greater than 100μL 2% FCS. Analysis of human and AML engraftment from NOD/SCID xenotransplant model was carried out by flow cytometry.

HARVEST OF MURINE BM

Twelve weeks following transplantation, mice were sacrificed by cervical dislocation. Femurs, tibias and pelvis were dissected from the mice and stored in RT PBS. Bones were flushed into 1mL of RT PBS in a 5mL snap-top polystyrene tube (BD Pharmingen, San Diego). Both ends of each bone were removed to provide an opening and an insulin syringe (microfine (29G), BD Pharmingen, San Diego) containing 1mL of RT PBS was inserted into one end of each bone and the lumen of the bone was washed with medium pressure. Flushing was repeated twice for both ends of the bone until the bone appeared relatively white.

Red blood cells were lysed by cooling the cell suspension solution on ice for 5 minutes and then adding 3mLs of cold ammonium chloride solution (Stem Cell Technologies, Vancouver, BC, Canada) for 5 minutes at 4°C. The lysis reaction was stopped by adding 0.5mLs of FCS. Cells were spun down and resuspended in cold 2% FCS. Cells were stored on ice until ready for antibody labeling.

ANTIBODY LABELLING

A mix of human specific FITC-conjugated anti-CD19, PE-conjugated anti-CD33 and PerCP-conjugated anti-CD45 antibodies was prepared (5μL / sample / antibody for all stains and compensation/isotype controls; all antibodies from BD Pharmingen, San Diego). Also prepared were FITC, PE and PerCP single color compensation control tubes (5mL snap-top polystyrene as before) and a combined FITC/PE/PerCP matched isotype control tube. Fifteen μL of the antibody mix was aliquoted into each tube of a fresh set of tubes for antibody labeling. Forty μL of each cell suspension was dispensed into the appropriate antibody labeling tube and left to label for 30 minutes at 4°C. Cells were washed in 2mL 2% FCS and resuspended in 500μL of 2% FCS supplemented with a cell impermeant DNA dye for live/dead discrimination, either 100ng/mL 4,6-diamidino-2-phenylindoiole (DAPI, UV excited, Sigma-Aldrich, St. Louis) or TOPRO-3 (HeNe (633nm) excited, Molecular Probes Inc., Eugene, OR).

FLOW CYTOMETRY

Flow cytometry analysis was conducted on a BD LSR II (BD Bioscience, San Jose, USA) using CellQuest software (BD Biosciences). Subsequent analysis was preformed with FlowJo software (Ashland USA). Lasers and bandpass filters required for analysis are listed in **Table 2.1.**

During analysis, the photomultiplier gains were set so that the background signal from the combined isotype control gave 1-5% positive cells in each collection channel. Compensation amount was set according the detected spectral overlap. To analyse the engraftment, 4 dotplots were drawn as in A, B, C and D in **Figure 2.4** (440/40nm vs sidescatter (SSC), forward scatter (FSC) vs SSC, 695/40nm vs SSC and 530/30nm vs 575/26nm). Dead cells

Laser	Bandpass Filter	Fluorochrome
488nm (blue)	440/40	DAPI
	530/30	FITC
	575/26	PE
	660/20	Topro-3
	695/40	PerCP
	780/60	PE-Cy7
633nm (red)	660/20	APC

Table 2.1: Lasers and bandpass filters required for flow cytometry analysis on engraftment and sorting of cells

were excluded from the analysis via a region (R1) around the live, unstained cells as in **Figure 2.4A**. These cells were displayed on a FSC vs SSC plot and the lymphoid and myeloid cells were selected for further analysis but debris was excluded via a region (R2) as in **Figure 2.4B**. Cells that fell into the first 2 regions were displayed on a 695/40nm vs SSC plot and generous region around the CD45-PerCP positive cells was drawn as in **Figure 2.4C** (R3). These $CD45^+$ cells were displayed on a CD19-FITC vs CD33-PE (530/30nm vs 575/26nm) dotplot and quadrants were drawn to define $FITC^+/PE^-$ cells and $FITC^-/PE^+$ cell subsets as in **Figure 2.4D**. The scatter characteristics of cells was confirmed as consistent with myeloid (high FSC and SSC, example in **Figure 2.4E**) and lymphoid (low FSC and SSC, example in **Figure 2.4F** (in this example there are no $CD19^+$ cells)). Engraftment was classed as myeloid leukaemia if a population of $CD45^+/CD33^+$ cells was present without an accompanying $CD45^+/CD19^+/CD33^-$ cell population.

For analysis of human cell engraftment (as in virus transduced CB MNC), bone marrow was harvested and analysed in the same manner as for AML engraftment. **Figure 2.5** gives an example of the dot plots for analysis, layed out in the same manner as described in **Figure 2.4**. Lymphoid engraftment is apparent in this sample (**Figure 2.5D**).

SORTING OF LEUKAEMIA SAMPLES INTO STEM CELL AND EARLY PROGENITOR POPULATIONS

PREPARATION OF SAMPLES AND ANTIBODY LABELING

Leukaemia samples were removed from liquid nitrogen, rapidly thawed and washed in 50mL of ice cold 2% FSC. Cells were spun down at 4°C, resuspended in 1mL of 2% FSC and counted. Cells were passed through a

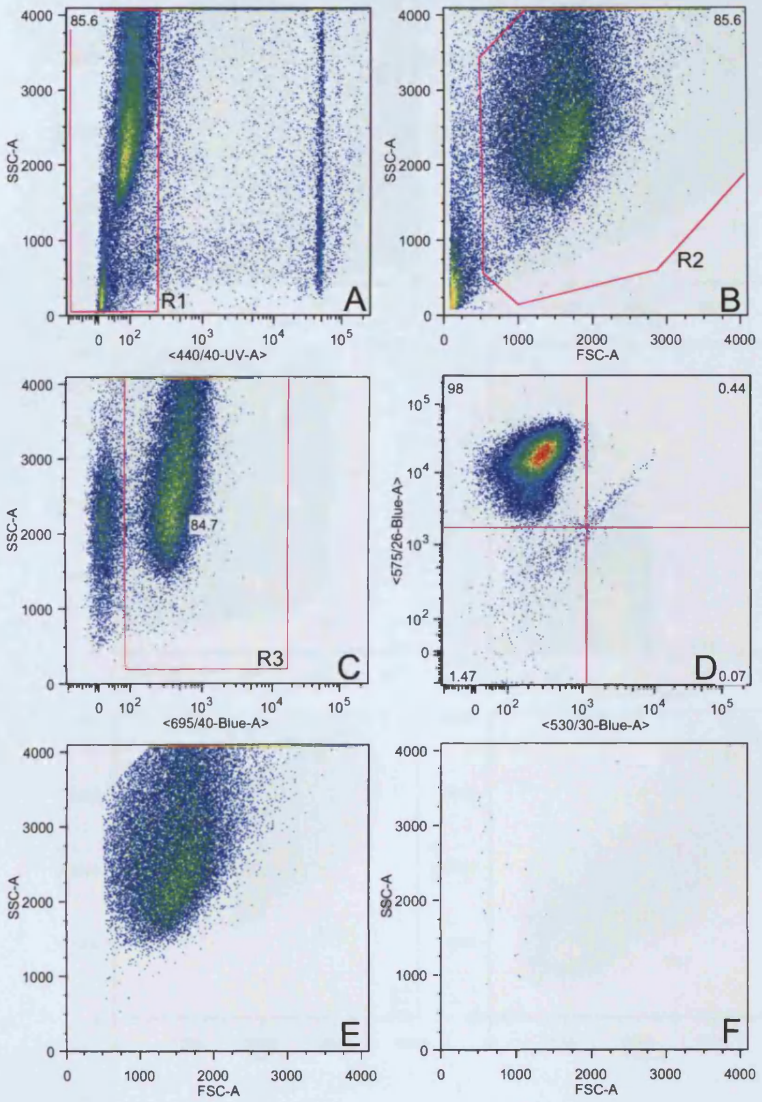


Figure 2.4: Analysis of AML engraftment in murine xenotransplant model

See text for description of figure.

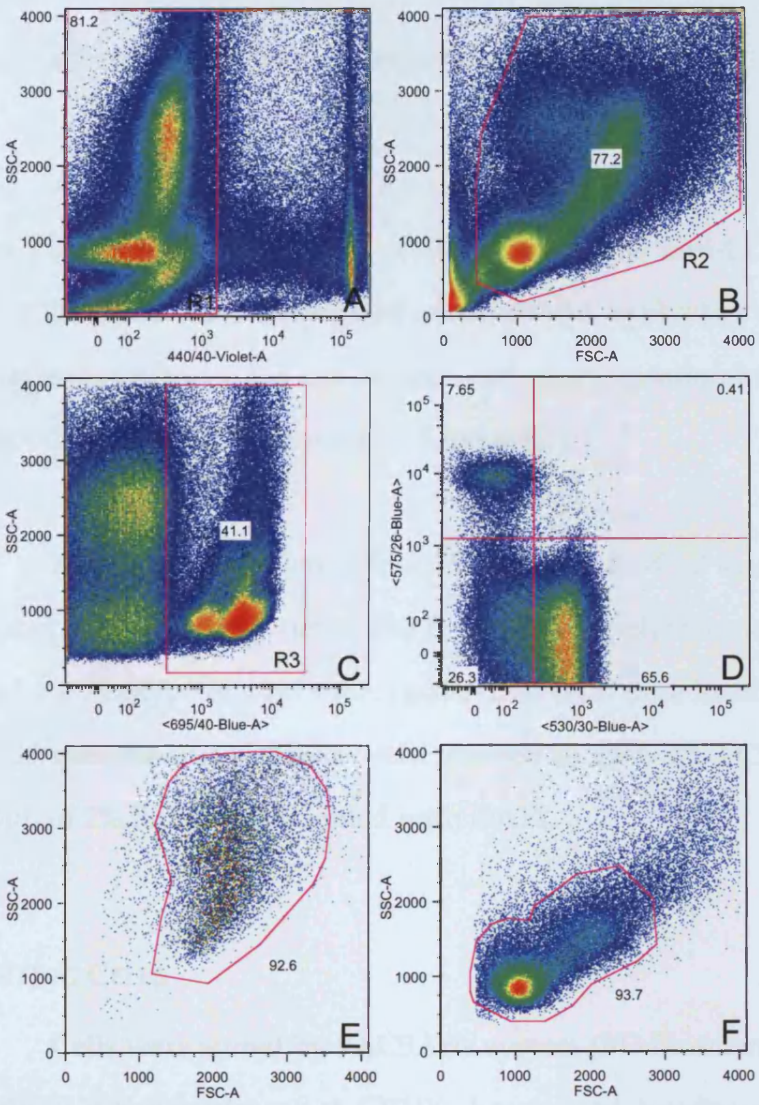


Figure 2.5: Analysis of human cell engraftment in murine xenotransplant model

See text for description of figure.

70μm cell strainer (Falcon) to remove any clumps.

A mix of human specific FITC-conjugated Lineage cocktail, PerCP-conjugated anti-CD34, PE-Cy7-conjugate anti-CD38, PE-conjugated anti-CD123 and APC-conjugated anti-CD45RA antibodies was prepared (5μL / sample / antibody for all stains and compensation/isotype controls; all antibodies from BD Pharmingen, San Diego).

Also prepared were FITC, PE, PerCP, PE-Cy7 and APC single color compensation control tubes (5mL snap-top polystyrene as before) and a combined isotype control tube. Leukaemia cells were incubated with antibody for 30 minutes at 4°C. Cells were washed in 2mL 2% FCS and resuspend in 500μL of 2% FCS supplemented with DAPI.

SORTING CELLS

Cells were sorted by FACSAria system (BD Biosciences) with assistance of the FACS department at CRUK. Lasers and bandpass filters required to analyse this set of fluorochromes are listed in **Table 2.1**. Compensation and isotype controls were set up as described above. Cells were sorted into three subsets that correspond to distinct compartments in non-malignant haematopoiesis, HSC (CD34+/38-), CMP and GMP (**Figure 2.6**). Populations were sorted based on the lack of lineage markers in all cases, the stem cells express CD34 cell surface marker, but not CD38 (Lin-/CD34+/CD38-). The CMPs and GMPs express CD34, CD38, and CD123 and are distinguished by presence (GMP) or absence (CMP) of CD45RA. Cells were sorted into FCS and then recovered by centrifugation.

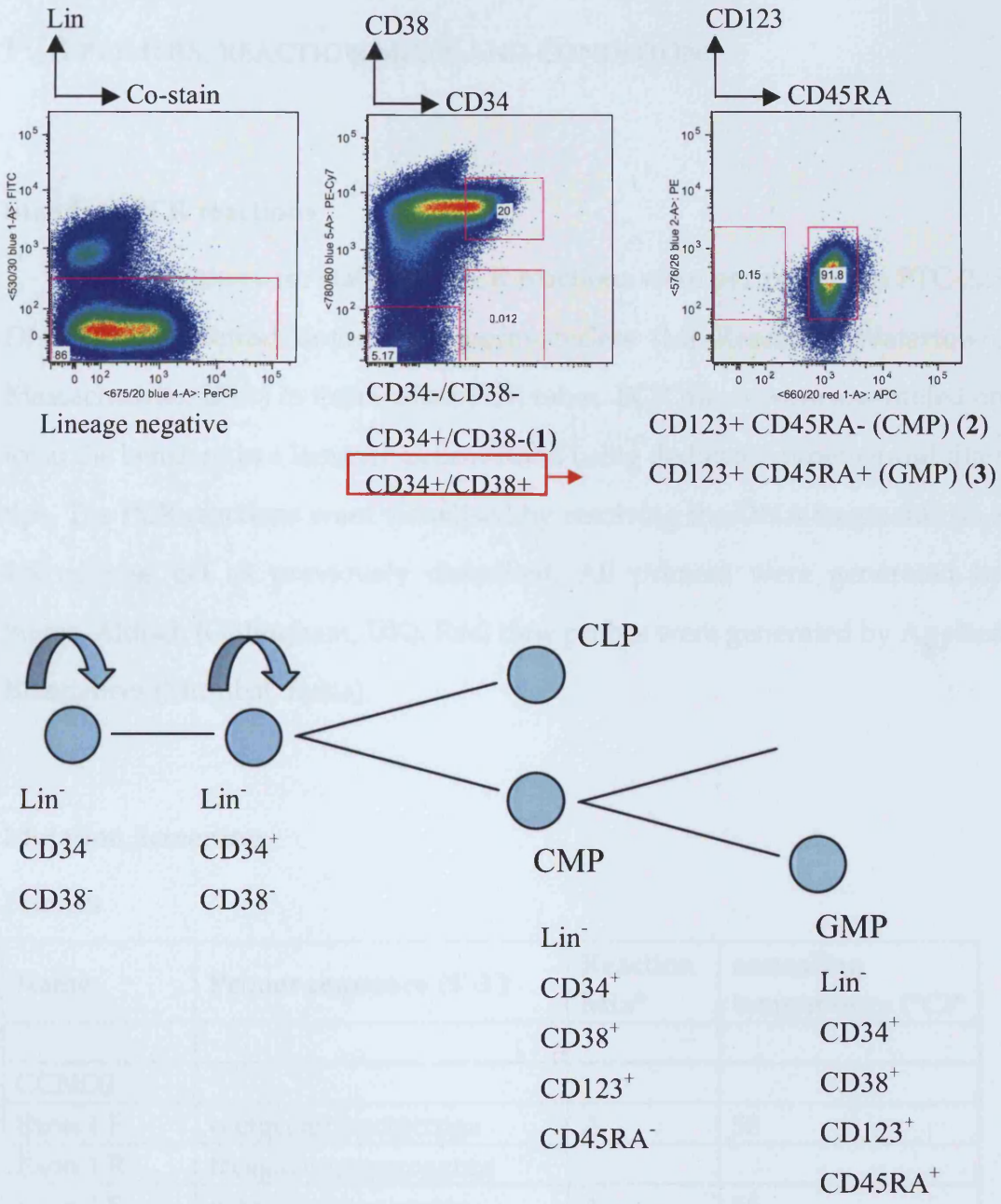


Figure 2.6: Sorting leukaemia samples into subsets that correspond to stem and early progenitor population

Leukaemia samples are sorted based on the presence or absence of cell surface markers. The leukaemia sample is stained with fluorescent conjugated antibodies, including a cocktail of mature lineage markers, CD34, CD38, CD123 and CD45RA. HSCs are sorted for absence of mature lineage markers (Lin-), presence of CD34(+) and absence of CD38(-). CMP and GMP populations are Lin-/CD34+/CD38+/CD123+ with GMP expressing CD45RA(+) and CMP without.

PCR PRIMERS, REACTION MIXES AND CONDITIONS

Standard PCR reactions

Unless otherwise stated, all PCR reactions were performed on PTC-225 DNA Engine Tetrad automated thermocyclers (MJ Research, Watertown, Massachusetts, USA) in thin walled PCR tubes. PCR mixes were assembled on ice at the bench or in a laminar airflow hood using dedicated pipettes and filter tips. The PCR reactions were visualised by resolving the DNA fragments on a 1% agarose gel as previously described. All primers were generated by Sigma-Aldrich (Gillingham, UK). Real time probes were generated by Applied Biosciences (Mumbai, India).

Mutation Screening

Primers

Name	**Primer sequence (5'-3')**	**Reaction Mix***	**annealing temperature (°C)***
CCND3			
Exon 1 F	ccctgcctgttcgctgcccga	A	58
Exon 1 R	tccagacccgggagcggtgg		
Exon 2 F	gctccagacccagcagtga	A	55
Exon 2 R	gggagacaatagctgtcggg		
Exon 3 F	ccactgaccccacccttttc	A	55
Exon 3 R	gcaaaaatgtgcatagggctc		
Exon 4 F	ctcatggttccctttcctcct	A	57
Exon 4 R	atgaatggagaggctgctgc		
Exon 5 F	cactcttctcccatgttccca	A	52
Exon 5 R	actccagagggcctctccag		
CEBPA			
1F	tcgccatgccgggagaactctaac	B	60
1R	cctgctgccggctgtgctggaac		

2F	cttcaacgacgagttcctggccga	B	60
2R	agctgcttggcttcatcctcct		
3F	ccgctggtgatcaagcagga	B	60
3R	ccggtactcgttgctgttct		
4F	caaggccaagaagtcggtggaca	B	60
4R	cacggtctgggcaagcctcgagat		
FLT3			
Exon 14-15 F	gcaatttaggtatgaaagccagc	A	56
Exon 14-15 R	ctttcagcattttgacggcaacc		
Exon 20 F	ccaggaacgtgcttgtca	A	
Exon 20 R	tcaaaaatgcaccacagtgag		
PU.1			
Exon 1 F	tcacccagggctcctgtagctca	B	62
Exon 1 R	tcgtgggcaggcaggcaggcgtcc		
Exon 2 F	tgatggggaccagcgtgcggggt	B	62
Exon 2 R	tctctccagaccccaggaccaggc		
Exon 3 F	actataacctttcctgccctgcc	B	62
Exon 3 R	agcctgtgtcagcttcctgtgaag		
Exon 4 F	tgcactccttctctccccagctgacc	B	62
Exon 4 R	acacacacgcgactcggtggcgtg		
Exon 5 F	ccgggcccctgtgcgtacgcaagg	B	62
Exon 5 R	ccgggagcgtcctccctgtgtccg		
RUNX1			
Exon 3 F	gcctgtcctcccaccaccctctc	B	60
Exon 3 R	agctgcttgctgaagatccg		
Exon 4 F	gtgggtttgttgccatgaaacg	B	58
Exon 4 R	catccctgatgtctgcatttgtcc		
Exon 5 F	cccaaggaatctgagacatggtcc	B	58
Exon 5 R	tgttcaggccaccaacctcattc		
WT1			
Exon 7 F	gacctacgtgaatgttcacatg	A	60
Exon 7 R	acaacacctggatcagacct		
Exon 8 F	cctttaatgagatccccttttcc	A	58
Exon 8 R	ggggaaatgtggggtgtttcc		
Exon 9 F	tgcagacattgcaggcatggcagg	A	58
Exon 9 R	gcactattccttctctcaactgag		
Exon 10 F	acttcactcgggccttgatag	A	50

Exon 10 R	gtggagagtcatacttgaaag		

*Reaction mixes are listed immediate beneath. Annealing temperature listed takes place of x temperature in listed programs below.

Reaction Mixes

Mix A

Reagents	Concentration	Volume
Template	50-100 ng/μL	2 μL
Primer 1	20 μM	0.05 μL
Primer 2	20 μM	0.05 μL
dNTPs	25mM each	4 μL
Buffer	10x	2.5 μL
Pic Taq	5U/ μL	0.25 μL
dH_2O		To 25 μL

10x buffer (Promega, Madison, USA), Taq (Perkin Elmer, Foster City, CA; Cancer Research, London, UK), dNTPs (Cancer Research UK, London UK)

Mix B

Reagents	Concentration	Volume
Template	50-100 ng/μL	2 μL
Primer 1	20 μM	0.5 μL
Primer 2	20 μM	0.5 μL
dNTPs	25mM each	2 μL
Buffer	10x	2.5 μL
$MgCl_2$	25 mM	1.25 μL
DMSO	stock	1.25 μL
Taq polymerase	5U/ μL	0.1 μL
dH_2O		To 25 μL

Taq and 10x Buffer from Invitrogen (Paisley, UK)

PCR Programs Mutation Screening

CCND3

Step	Temp (°C)	Time (min)
1	96	2:00
2	96	0:30
3	x	0:30
4	72	0:30
5	Go to 2, 34 times	

6	72	10:00
7	4	Hold
8	End	

Annealing temperatures (x) are given in primer chart above.

CEBPA

Step	Temp (°C)	Time (min)
1	95	2:00
2	95	1:00
3	60	0:40
4	72	1:00
5	Go to 2, 39 times	
6	72	10:00
7	4	Hold
8	End	

FLT3

Step	Temp (°C)	Time (min)
1	94	3:00
2	94	0:30
3	56	1:00
4	72	2:00
5	Go to 2, 34 times	
6	72	10:00
7	4	Hold
8	End	

PU.1

Step	Temp (°C)	Time (min)
1	95	5:00
2	95	0:40
3	62	0:40
4	72	1:30
5	Go to 2, 34 times	
6	72	10:00
7	4	Hold
8	End	

RUNX1

Step	Temp (°C)	Time (min)
1	94	2:00
2	94	0:30
3	x	0:30
4	72	0:30
5	Go to 2, 34 times	
6	72	10:00
7	4	Hold
8	End	

Annealing temperatures (x) are given in primer chart above.

WT1

Step	Temp (°C)	Time (min)
1	92	2:00
2	92	0:45
3	x	1:00
4	72	1:30
5	Go to 2, 29 times	
6	72	10:00
7	4	Hold
8	End	

Annealing temperatures (x) are given in primer chart above.

WAVE analysis for FES mutation screening

Reagents	Concentration	Volume
Template	10 ng/μL	2 μL
Primer 1	100 pmol/ μL	0.05 μL
Primer 2	100 pmol/ μL	0.05 μL
dNTPs	100x	0.1 μL
Amplitaq buffer II	10x	1 μL
$MgCl_2$	25 mM	1.2 μL
Amplitaq Gold		0.1 μL
dH_2O		To 10 μL

Amplitaq Gold, buffer and dNTPs from Applied Biosystems (Foster City, USA)

Primers

Exon	Primer sequence	PCR conditions	DHPLC analysis temperature (°C)
13	tggaggacacagcccttcta	AT66	63, 66
	acaagtggggaatcagatgg		
14	gggtaggaagcggagaagag	touchdown	63, 66
	ctgtgccatgcagtgtaacc		
15 + 16	ccaagctcttctcccatcat	touchdown	63, 68
	gggccctaggggtataggta		
17	cagcttttgtccttggcttt	touchdown	61, 64
	catggagtccagggtagaca		
18	actcacgctgcctcacct	AT66	62, 65
	gaaggttggcattaagggagt		
19	tcatggtatcccccagagtc	touchdown	63, 66
	tgtcagcatatgagctggag		

AT66 (*FES*)

Step	Temp (°C)	Time (min)
1	95	12:00
2	95	0:30
3	66	0:30
4	72	0:30
5	Go to 2, 35 times	
6	72	10:00
7	15	Hold
8	End	

Touchdown (*FES*)

Step	Temp (°C)	Time (min)
1	95	12:00
2	95	0:30
3	60	0:30
4	72	1:00
5	Go to 2 4 times	
6	95	0:30
7	60-0.5/c	0:30
8	72	1:00
9	Go to 6 19 times	
10	95	0:30
11	50	0:30

12	72	1:00
13	Go to 10, 15 times	
14	72	10:00
15	15	hold
16	End	

Sequencing

Template and Reaction Concentrations

Template	**Quantity used per reaction (ng)**
Double Stranded	150-200
PCR Templates:	
100-200 bp	1-3
200-500 bp	3-10
500-1000 bp	5-20
1000-2000 bp	10-40

Reaction Mix

Reagents	**Amount**
Template	(see above)
Primer	3.2 pmol
Terminator Ready Reaction Mix*	8.0 μL
dH_2O	To 20 μL

*used at stock concentration for slab gels and 1:16 dilution for capillary gels
*Big Dye Terminator v3.1 (PE Applied Biosystems, Foster City, CA, USA)

Sequencing PCR Program

Step	**Temp (°C)**	**Time (min)**
1	96	1:00
2	96	0:10
3	50	0:05
4	60	4:00
5	Go to 2, 24 times	
6	4	hold

Real Time PCR reaction

Standard Real-time PCR

Reagents	Amount (25μL reaction)
DNA	varies
TaqMan Universal PCR master mix	12.5μL
Forward primer	300nM
Reverse primer	300nM
Probe	200nM

(Master Mix from Applied Biosystems, Foster City, CA, USA)

For visualization

Reagents	Amount (25μL reaction)
DNA (100ng/μL)	5μL
TaqMan Universal PCR master mix - UNG	12.5μL
Forward primer (100ng/μL)	5μL
Reverse primer (100ng/μL)	5μL

(Master Mix from Applied Biosystems, Foster City, CA, USA)

All the reactions were carried out in duplicate with the following conditions: an initial 50°C for 2 minutes and 95°C for 10 minutes, followed by 40 cycles of a two stage PCR were used, consisting of 95°C for 15 seconds and 60°C for one minute. β2M was used as an endogenous reference gene (forward primer, 5'ggaattgatttgggagagcatc; reverse primer, 5'caggtcctggctctacaatttactaa; probe, 5'agtgtgactgggcagatcatccaccttc). Analysis was carried out using the software provided with the ABI Prism 7700.

Nested PCR for *CEBPA* cloning

Using TaKaRa LA Taq (taq, buffer, dNTPs) kit (Takara Bio Inc, Shiga, Japan)

Reagents	Concentration	Volume
Template[a]	50-100 ng/μL	2.5 μL
Forward Primer[b]	10 pmol/ μL	0.5 μL

Reverse Primer[b]	10 pmol/ μL	0.5 μL
dNTP mixture	2.5 mM each	4 μL
LA PCR Buffer II	10x	2.5 μL
LA Taq	5 units/μL	0.25 μL
dH_2O		To 25 μL

a: First round genomic DNA with a concentration of 50-100ng/μL is used. Second round, PCR product of first round PCR is used as template
b: First round CHR19P1 (5' – TTGCCCAGATGAAACTGCTTCTTTACTGCG - 3') and CHR19P4 (5' – CTGGAATTAGCACTGAACTCAGAGGGTTTG - 3').
Second round *CEBPA* 1F and *CEBPA* 4R (as above).

Round 1

Step	Temp (°C)	Time (min)
1	95	1:00
2	100	0:05
3	98	0:05
4	60	1:00
5	62	11:00
6	Go to 2, 34 times	
7	72	10:00
8	4	hold

Round 2

Step	Temp (°C)	Time (min)
1	95	1:00
2	100	0:05
3	98	0:05
4	60	1:00
5	62	3:00
6	Go to 2, 34 times	
7	72	10:00
8	4	hold

SOLUTIONS AND RECIPES

Analytical or equivalent grade chemicals were obtained from Sigma, UK or BDH, UK. All buffers and solutions used de-ionised water as the diluent and were sterilised by autoclaving or filtration as appropriate.

Electrophoresis

10x TBE electrophoresis buffer (1L)

108g Tris base

55g boric acid

40mL 0.5 M EDTA, pH 8.0

Working stock = 1x

Agarose gel electrophoresis loading buffer (10x)

25mM EDTA

50% glycerol(v/v)

0.25% bromophenol blue (w/v).

Slab Sequencing Gel (50mL)

5mL 10x TBE

45mL Acrylamide Mix

250μL 10% APS

30μL TEMED

Western Blotting

Cell Lysis Solution (10mL)

10mL CelLytic-M Mammalian Cell Lysis/Extraction Reagent (Sigma)

1 Complete Mini, EDTA-free Protease Inhibitor Cocktail Tablet (Roche Diagnsotics)

12% Resolving Gel (20mL)

6.6mL dH_2O

8.0mL 30% (w/v) acrylamide/0.8% bisacrylamide (UltraPure Protogel. 37.5:1; National Diagnosis)

5.0mL 1.5M Tris (pH 8.8) (45.4g/250mL Trizma base, pH with HCl)

0.2mL 10% SDS (w/v 50g/100mL)

0.2mL 10% APS (w/v 1.5g/100mL, prepare fresh)

8.0µL TEMED

5% Stacking Gel (4mL)

2.7mL dH_2O

670µL 30% (w/v) acrylamide/0.8% bisacrylamide

500µL 1.0M Tris (pH 6.8) (30.3g/250mL Trizma base, pH with HCl)

40µL 10% SDS

40µL 10% APS

4.0µL TEMED

6x Loading Buffer

0.225M TrisCL (pH 6.8)

50% glycerol

5% SDS

0.05% bromophenol blue

0.25M DTT

10x Running Buffer (1L)

30g Trizma base

144g Glycine

10g SDS

pH to 8.5

working stock = 1x

Transfer Buffer (2L)

22.5g glycine

4.84g Trizma base

400mL methanol

10x TBS (1L)

24.2g Trizma base

80g sodium chloride

pH to 7.6

working stock = 1x

TBST (2L)

200mL 10x TBS

10mL 10% Tween

Blocking Solution

5% marvel milk in TBST

Primary Antibody Solution

5% BSA in TBST

Secondary Antibody Solution

5% marvel milk in TBST

Bacteria Culture

LB-Agar

LB-Broth 1 litre (CRUK)

15g bacto-agar

Melt then cool to 50°C prior to addition of antibiotics. Pour into petri dishes and leave to set at room temperature in laminar airflow hood. Dry plates in 37°C incubator 2-3 hours before use.

X-gal

Stock solution 20mg/ml. 0.2g dissolved in 10ml of N,N-Dimethylformamide. Filter, sterilise through a 0.4μm filter and store at –20 °C.

Ampicillin

Stock solution 50mg/ml. Filter, sterilise through a 0.4μm filter and store at –20°C. Use at a final concentration of 0.1mg/ml in media.

Kanamycin

Stock solution 50mg/ml. Filter, sterilise through a 0.4μm filter and store at –20°C. Use at a final concentration of 0.05mg/ml in media.

Cell Culture

MC colony lysis buffer (20mL)

910μL Protease K (20mg/ml) (must be added fresh)

100μL Tween 20

100μL NP40

800μL 250mM Tris pH 7.5

18.09mL dH_2O

Miscellaneous

TE buffer pH 7.4 or pH 8.0

10mM Tris-HCl (pH titrated with 10M NaOH)

1mM EDTA pH 8.0

Digestion Buffer for DNA extraction

100mM NaCl

10mM Tris.Cl, pH 8

25mM EDTA, pH 8

0.5% SDS

SSCP loading buffer

4.75ml formamide

125µl 1% bromophenol blue

125µl 1% xylene cyanol

CHAPTER 3: OCCURRENCE AND PATTERNS OF GENE-SPECIFIC MUTATIONS IN NORMAL KARYOTYPE ACUTE MYELOID LEUKAEMIA

BRIEF INTRODUCTION AND DESCRIPTION OF STUDY

Cytogenetic analysis is important for the clinical management and prediciting prognosis in AML[140-142]. However, approximately half of all AMLs have normal karyotype. The lack of overt cytogenetic markers prevents risk stratification or minimal residual disease monitoring in these patients. These NK patients form a heterogeneous disease group with variable responses to therapy. They are classified as intermediate risk due to having worse outcome than those with favourable prognosis, but better than those with poor prognosis. While NK AML may appear to be the same on the chromosomal level, there is significant variation in this group at the molecular level. This molecular variability has been substantiated over the years with the discovery of gene-specific mutations and changes in gene expression levels that affect the clinical outcome of NK AML patients[70]. We are reaching a point therefore where the characterisation of mutations should be categorised at diagnosis in a similar manner to cytogenetic testing which has become part of the standard clinical routine in treatment of AML.

Aside from these clinical applications, documenting the mutations

which are present in NK AML can provide novel insights into the process of leukaemogenesis.

In this chapter, we have set out to examine the patterns and relationships between gene-specific mutations in NK AML. To accomplish this, we performed mutation analysis on a panel of 88 NK AML patient samples for genes known to be mutated in AML.

Examining Mutation Status of Normal Karyotype AML

Patterns of mutations in NK AML

All samples screened were collected and analysed in accordance to the Local Research Ethics Committee. These 88 patients consisted of a cross section of age and FAB types. APL was excluded from this study due to the unique nature of this leukaemia. All other FAB types were represented except for M0 (**Table 3.1**). Eighty-one of the samples were taken at diagnosis, 7 were taken at relapse. Samples were of peripheral blood (PB) in 55 cases, bone marrow (BM) in 13 cases and leucopheresis product in the remaining 20. All samples had undergone density centrifugation for mononuclear cell enrichment prior to cyropreservation.

Initially, to validate the patient panel, six genes previously implicated in AML were screened for mutations. Comparing the mutation frequencies observed for these genes in our panel to published frequencies would indicate

Total		88
Gender % M:F		51:49
Age (years) median		53
FAB type	M0	-
	M1	35
	M2	22
	M4	19
	M5	10
	M6	2

Table 3.1: Characteristics of the normal karyotype panel

A panel of 88 patients with normal karyotype AML was validated by screening for genes known to be involved in leukaemogenesis. All FAB types were represented except for M0 and M3. This panel was subsequently used for prospective screening of novel gene involvement in AML.

whether our panel and methodology was robust enough to be used to search for novel gene mutations.

The initial six genes screened were *FLT3* (exons 14-15 and 20), *CEBPA* (entire coding region), *RUNX1* (exons 1-5), *NRAS* (exons 1 and 2), *cKIT* (exons 8 and 17) and *PTPN11* (exons 3 and 13). PCR for each exon indicated were carried out using DNA recovered from patient samples as template. Mutation detection was undertaken by single strand conformational polymorphism (SSCP). The amplified exons from patients were denatured to single strand fragments and then run through thin acrylamide gels. Under certain conditions, single stranded nucleic acids form secondary structure in solution. The secondary structure depends on the base composition and may be altered by a single nucleotide substitution. These substitutions result in a change in electrophoretic mobility under non-denaturing conditions. Fragments were detected by silver staining. When unusual banding patterns were observed the samples were directly sequenced for the gene and exon in question.

A total of 66 mutations in 56 patients were observed. Forty-seven patients had a single mutation (53%), 8 patients had two mutations (9%), one patient had 3 mutations (1%) and no mutation was observed in 32 patients (36%). The frequency of mutated genes was comparable to previously published screening studies[143], validating this panel for further mutation screening.

More recently 2 additional genes, *NPM1* (exon 12) and *WT1* (exons 2-10)

were investigated using a PCR direct sequencing approach to further refine the mutation profile of this cohort of patients. The addition of these genes to the screening panel increased the percentage of normal karyotype patients with a detectable mutation from 64% to 85% (75/88), with 54 mutations in *NPM1* and 7 mutations in *WT1* being added to the panel (**Figure 3.1**).

This panel now provides a more complete overview of the types of mutations that occur and the associations between these mutations in NK AML. Nearly half of the patients screened contained mutations in more than one gene (45%), supporting the idea that multiple genetic events are required for the onset of leukaemia (**Figure 3.2**). In our panel of 88 patients, 35 patients had a single mutation (40%), 28 patients had two mutations (32%), 11 patients had three mutations (12.5%) and one patient had four mutations (1%).

According to Gilliland's two-hit hypothesis, mutations affecting proliferation and self-renewal capacity (Class I) in association with mutations causing blocks in differentiation (Class II) are required for overt leukaemogenesis[86]. Based on the published functional properties of these gene products, we classified *FLT3, NRAS, KRAS* mutations as Class I and *CEBPA, RUNXI, NPM1, WT1* mutations as Class II. In our panel of NK AMLs, 44% (39/88) of patients have mutations in genes from each mutation class, 32% (28/88) have Class II mutations alone and 9% (8/88) had mutations restricted to Class I mutations (**Figure 3.3**).

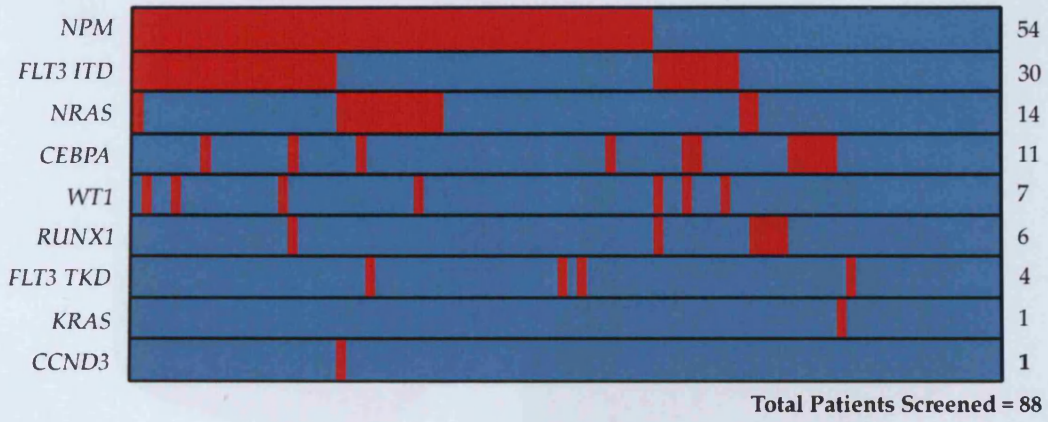


Figure 3.1: Gene specific mutations in Normal Karyotype AML

A panel of eighty-eight normal karyotype AML patients was screened for mutations in *NPM* (exon 12), *FLT3* (exons 14-15, ITD and 20, TKD), *CEBPA* (entire coding region), *WT1* (exons 2-10), *RUNX1* (exons 1-5), *CCND3* (exons 1-5), *NRAS* (exons 1 and 2), *KRAS* (exons 1 and 2), *cKIT* (exons 8 and 17), and *PTPN11* (exons 3 and 13) (no mutations were observed and therefore these genes were not included). Red bars indicate the presence of mutation, blue indicates a normal sequence for the particular gene.

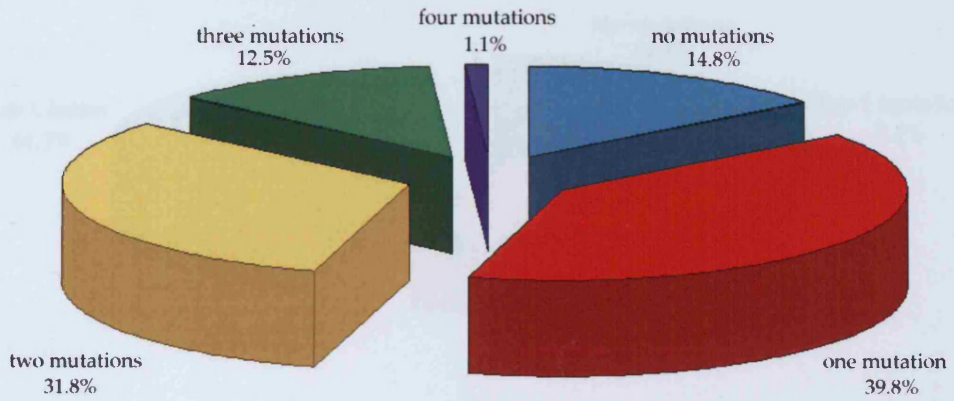


Figure 3.2: Most NK AMLs have mutations in more than one gene

A total of 128 mutations were identified in the 88 patients. Thirteen patients had no mutations in any of the genes screened, 35 patients had one mutation, 28 patients had two mutations, 11 patients had three mutations and one patient had four mutations.

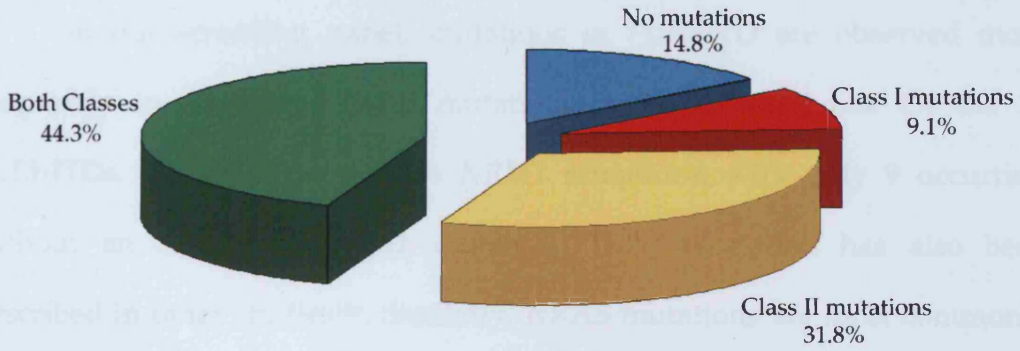


Figure 3.3: Class I and Class II mutations appear to cooperate in NK AML

It has been postulated that for leukaemia to arise, cooperating mutations that result in both the deregulation of proliferation (Class I) and block in differentiation (Class II) are required. In our study, 44.3% of patients screened had at least one mutation in each class.

MUTATIONS OCCUR IN A NON-RANDOM MANNER

In our screening panel, mutations in *FLT3*-ITD are observed more frequently in association with mutations in *NPM1* than not. Of the 30 *FLT3*-ITDs, 21 occur along with *NPM1* mutations, with only 9 occurring without an associated *NPM1* mutation. This association has also been described in other studies[144]. Similarly, *NRAS* mutations are most commonly seen in patients with coexisting *NPM1* mutations. Mutations in *FLT3* and *NRAS* are nearly always mutually exclusive, with only one patient of the 44 displaying mutations in both of these genes. The negative association of *NRAS* and *FLT3* mutations has also been reported in previous studies[68]. This mutual exclusion may indicate that *NRAS* and *FLT3* are individually sufficient for the deregulation of proliferation control required in leukaemogenesis.

MUTATION SCREENING OF NOVEL GENES

Although the majority of patients in our panel were found to have at least one mutation in the eight genes screened, 15% (13/88) had no detectable genetic lesion, indicating that novel mutations are likely to be present in these patients. Two further candidates, *CCND3* and *FES*, were selected for mutation analysis in our series of cases.

CYCLIN D3 (*CCND3*)

Cell cycle regulation is driven by the cyclin family of proteins, working in conjunction with cyclin-dependant kinases (CDKs) at various stages of the

cell cycle. Cyclin D3, located on 6p21, plays a crucial role in the positive regulation of cell proliferation. *CCND3* was chosen as a candidate gene for mutation screening because of its reported interaction and negative regulation of *RUNX1 (AML1)*[145, 146]. Cyclin D3 was identified to negatively regulate the transactivation activity of *RUNX1* in a dose-dependent manner in competition with *CBFβ*, leading to a decreased binding affinity of *RUNX1* for its target DNA sequence[145].

Cyclin D3 also interacts with other proteins that have been implicated in leukaemogenesis. The D type cyclins are regulated by the RAS/MEK/ERK pathway, with *RAS* an important mutational target in AML. Additionally, *CCND3* knockout mice die *in utero* at 14-15 days of gestation due to defective haematopoietic and neuronal differentiation demonstrating the impact of *CCND3* in normal haematopoiesis[147]. Cyclin D3 expression levels have also been used as a prognostic marker in other haematological malignancies such as follicular lymphoma, chronic lymphocytic leukaemia and large B-cell lymphoma[148-151]. In addition, it is targeted for translocation to various partner genes in mature B-cell malignancies and multiple myeloma[152, 153] and a variety of solid tumours[154-156].

Eighty-six normal karyotype AML samples were screened for mutations in all five exons of *CCND3* via SSCP analysis as described above (**Figure 3.4**). Samples which exhibited unusual SSCP traces were investigated for mutation by direct sequencing of PCR products. Thirty-five samples were selected for direct sequencing based on abnormal banding patterns. Direct

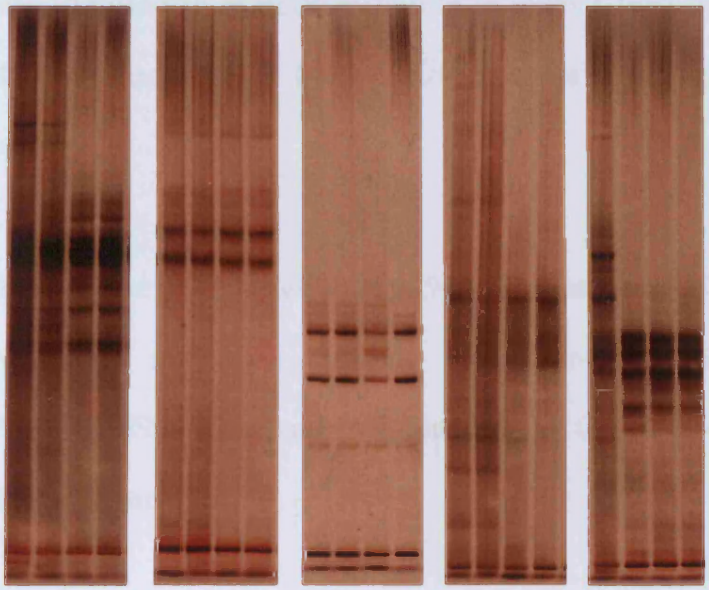


Figure 3.4: Examples of SSCP gels for *CCND3* screening

Single strand conformational polymorphism was used to screen all exons of *CCND3*. Each exon is amplified by PCR, the fragments then denatured and run on a thin acrylamide gel. Alterations in the DNA sequence caused by mutations can change the conformation of the fragment, causing it to migrate differently from wild type products. Bands were visualised by silver staining and traces which appear different from the majority were tested for mutations by direct sequence analysis.

sequencing was required to exclude mutation in exon 1 (n=7 tested), exon 3 (n=5), exon 4 (n=6) and exon 5 (n=16). No abnormal SSCP patterns were observed for exon 2.

A single in-frame 51 bp deletion of nt 940-990 (Genbank, NM_001760) in exon 5 resulted in the deletion of amino acids 259-275 (AQTSSSPAPKAPRGSSS) at the carboxyl-terminus of CCND3 was identified in a single patient (**Figure 3.5**).

Several sequence variations in exon 5 were identified and shown using the ensemble database to correspond to previously reported polymorphisms (www.ensembl.org). Six patients with polymorphisms affecting codons 258 and 259 were identified. Three patients contained a synonymous C > T point mutation at nucleotide 855, at codon 258 (A258A). This polymorphism occurred in conjunction with a T > G substitution at nucleotide 856 in 2 patients, which results in the substitution of an alanine for serine at codon 259 (S259A). The polymorphism alone was observed in one patient.

FELINE SARCOMA PROTO-ONCOGENE (FES)

The second gene tested in this series was *FES* (c-fes), it is located at 15q26.1 and encodes a 93 kDa non-receptor tyrosine kinase that is preferentially expressed in myeloid and endothelial cells, with expression restricted to mature haematopoietic cells of the granulocytic and monocytic lineages. *FES* is involved in cytokine receptor signal transduction, neutrophil

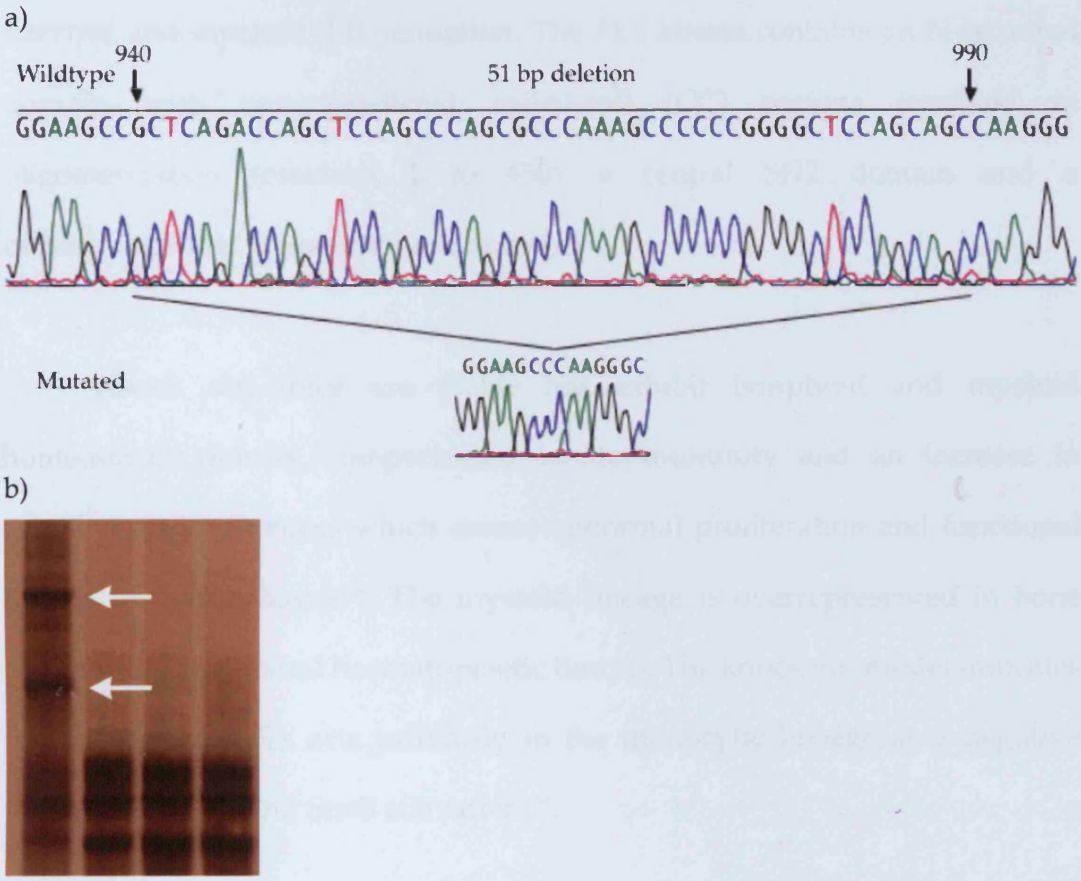


Figure 3.5: Sequencing trace of wild-type and mutated *CCND3*

a) Sequencing trace of wild-type and mutated *CCND3*. The wild-type sequence of exon 5 is demonstrated in the top panel with the corresponding sequencing trace from a cloned product showing the 51-bp deletion of nt 940-990 (Genbank, NM_001760) in the bottom panel. This mutation results in the deletion of amino acids 259-275 (AQTSSSPAPKAPRGSSS).

b) The SSCP trace from this patient sample along side 3 wild type samples, with the miscellaneous bands indicated by arrows.

survival and myeloid differentiation. The *FES* kinase contains an N-terminal domain with two predicted coiled-coil (CC) regions involved in oligomerization (residues 1 to 450), a central SH2 domain and a carboxy-terminal tyrosine kinase domain[157].

Knock out mice are viable but exhibit lymphoid and myeloid homeostastic defects, compromised innate immunity and an increase in cytokine responsiveness which causes abnormal proliferation and functional defects in macrophages[158]. The myeloid lineage is overrepresented in bone marrow and peripheral haematopoietic tissues. The knock out model indicates that endogenous *Fes* acts primarily in the monocytic lineage as a negative regulator of Stat3 and Stat5 activation[159].

As part of a larger mutational analysis study of the tyrosine kinome by the Vogelstein group[160] point mutations in exon 17 and 18 of the tyrosine kinase domain of FES were identified in patients with colorectal cancer. Elevated expression of *FES* has been observed in AML[161]. In haematopoeisis, *FES* induces terminal macrophage differentiation of myeloid progenitors and activation of myeloid transcription factor *PU.1*[162]. *FES* also suppresses *BCR-ABL* driven cell outgrowth[159]. Increased levels of *FES* expression results in a reduction of *BCR-ABL* transcription but this is dependant on *FES* retaining its functional kinase activity. *FES* was therefore selected as a candidate gene based on its link to malignancy and its involvement in the myeloid branch of haematopoiesis.

Samples from a mutation panel that included patients with favourable and poor risk cytogenetics (n=167) were screened for mutations in exons 13-19 of *FES*. Exons 13-19 encompass the tyrosine kinase domain of the protein. Mutation screening was conducted by the WAVE mutation screening method. Exons were individually amplified from genomic DNA by PCR. The amplified fragments were then denatured and annealed slowly to allow heteroduplex formation (**Figure 3.6**). Heteroduplexes form due to mismatches of base pairs from point mutations and elute more slowly from the column as compared to homoduplexes.

Chromatograms that differed from the majority of the samples were studied by sequencing of the samples in question. Direct sequencing was undertaken for exon 14 (n=3), exons 15 + 16 (n=12), exon 17 (n=6), exon 18 (n=5) and exon 19 (n=2). No abnormal chromatograms were observed for exon 13.

No mutations were detected. As with the case with *CCND3*, several synonomous polymorphisms were observed. In exon 14, GCG > GCA at nt 1839 coding for amino acid Ala, was observed (Genbank, NM_002005). In exon 16, GGC > GGA at nt 2001 coding for Gly was observed (**Figure 3.7**).

SUMMARY

The current study adds to the growing pool of information on mutation events in AML and examines the interactions between specific mutations in NK AML in patients attending St. Bartholomew's Hospital. To achieve this

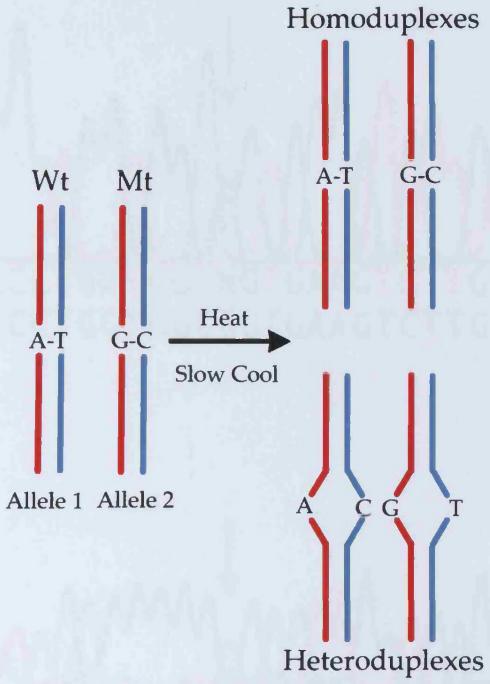
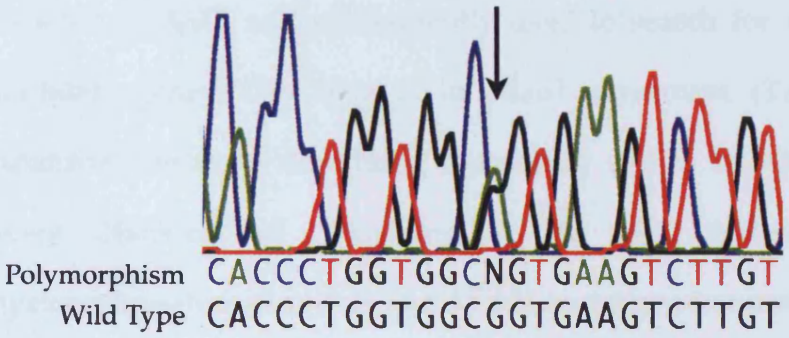


Figure 3.6: Wave mutation screening relies on heteroduplex formation

Exons of interest are amplified by PCR. The subsequent fragments are heated to denature the double stranded structure and reannealed slowly. If heterozygous point mutations are present, this allows the generation of heteroduplexes which will elute differently from a homoduplex when run through a dHPLC column.

a)



b)

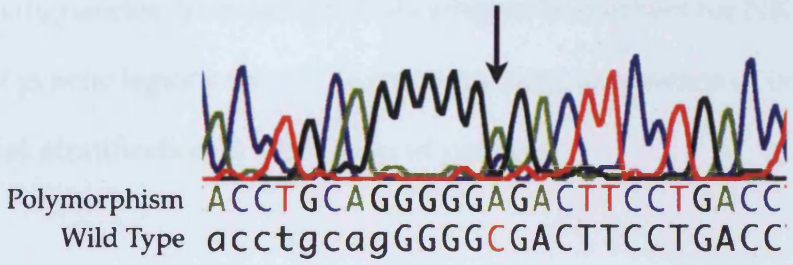


Figure 3.7: Polymorphisms detected in *FES*

No mutations were detected in the tyrosine kinase region of *FES*. However the wave mutation screening protocol was sufficient sensitive to detect polymorphisms within this region of the gene. a) Polymorphism in exon 14 G > A at nt 1839 (Ala) (Genbank, NM_002005); b) Polymorphism in exon 16 C > A at nt 2001 (Gly)

goal, a panel of NK AML patient samples was screened for mutations in genes involved in AML and subsequently used to search for novel mutations in candidate genes. Our data is in board agreement (**Table 3.2**) with the expansive literature describing mutational events in AML. Moreover, the recent discovery of mutations in the *JAK2* tyrosine kinase[163-165] in myeloproliferative disorders and *NPM1* mutations in normal karyotype AML [166, 167] as well as our group's study of *WT1*[168] are examples of the continuing discovery of novel mutations and their involvement in haematological malignancies. In the absence of cytogenetic markers for NK AML, the profiling of genetic legions should be recommended as a means of improving molecular risk stratification in this group of patients.

METHODS OF MUTATION SCREENING

Since mutation screening can be useful from a prognostic and therapeutic point of view, the development of high throughput techniques for screening remains important so that individual genetic profiles can be developed for patients from diagnosis. Moreover as AML is a heterogeneous disease with many patients doing poorly with current treatments, the tailoring of therapies towards specific genetic mutations may be the key to improving overall survival in this disease.

Besides clinical applications, high throughput screening methods are useful for discovering novel gene mutations, since large numbers of patient samples and individual gene exons must be screened to pick up rare genetic

	Our Frequency	Literature	Reference
NPM1	61%	50-60%	[63]
FLT3-ITD	34%	28-38%	[169]
NRAS	16%	13%	[170]
CEBPA	12.5%	4-15%	[169]
RUNX1	7%	7%	[171]

Table 3.2: Mutation frequencies detected in our panel are similar to those reported in literature

The mutation frequencies of genes in our panel were similar to those reported in literature, indicating that the panel and screening techniques are sufficiently sensitive/reproducable to be used to search for novel gene mutations.

events. Over time, more efficient and sensitive techniques have been developed for mutation detection, primarily based on separation of nucleic acids via molecular weights by electrophoresis[172]. Various electrophoresis techniques were utilised in our mutation screening protocols.

SSCP is able to detect heterozygous point mutations as well as more complex mutations such as insertions and deletions. Screening is faster with this method than direct sequencing, but mutation detection will vary depending on the size of product, its nucleotide sequence, the gel temperature, matrix and pH. The technique requires optimization by running at different conditions to increase the mutation detection rate[173], which is labour intensive compared to newer methods. Mutation detection rate also falls as the size of the amplicon increases. It is fair to assume therefore that the mutation detection rate of SSCP is below 100%. However, the mutation frequencies of the assessed genes in our study are similar to those in the literature, indicating that SSCP is adequate for routine laboratory investigations but would not be applicable in a diagnositic situation.

The possibility for high throughput analysis makes the dHPLC approach a valuable and altogether a more satisfactory tool in candidate gene screening. A disadvantage to this technique is that homozygous mutations are normally undetectable since mismatched heteroduplexes will not form. To overcome this problem, wild type DNA can be added to the test samples, thereby simulating heterogeneous mutation. Several studies have examined the sensitivity and specificity of dHPLC and with a detection rate of between

92-100% mutation[174-176].

Although direct sequencing was used in this study to screen *WT1* and in my later work relating *WT1* mutations and aUPD (Chapter 4), our laboratory now utilises capillary fragment analysis for screening of this gene. Capillary fragment analysis offers the advantage of being a high throughput technique that allows easy visualisation of the nature of the mutation by discrimination of the mutational events based on product size and provides a semi-quantitative measure of mutation load. However capillary fragment analysis is unable to detect point mutations[177]. All *WT1* mutations detected in our AML samples to date have been insertions or deletions, making this technique highly applicable for screening of this gene.

Ultimately, the goal is to be able to detect all mutations present in a panel of samples in a high-throughput and efficient manner. This ability is limited by the technology that is currently available. Confounding variables such as contamination with normal wild-type cells and clonal heterogeneity may make it difficult to pick up low levels of disease by electrophoretic techniques or direct sequencing technology.

COOPERATIVE MUTATIONS IN AML

By screening a panel of patients for a number of genes implicated in leukaemogenesis, distinct non-random patterns of mutations have now started to emerge. A large proportion of patients contained multiple mutations (45%)

and mutations representative of both classes (44%) (which affect self-renewal (Class I) and differentiation (Class II)), indicating that the association between mutation types may play an important role in leukaemogenesis in NK AML.

With the discovery of new gene associations in NK AML, functional testing can be preformed to determine whether the mutations are sufficient for the development of AML. Transduction of human HSCs with mutant forms of the gene or the creation animal models mimicking the mutation status observed in AML samples could be useful for determining whether these mutations are sufficent to cause the development of overt disease.

The persistence of AMLs with no mutations detected however, indicate the necessity to continue searching for novel gene mutations.

NOVEL GENE MUTATIONS IN NK AML

CCND3

The *CCND3* mutation detected in a single individual is the first reported case of this mutation in NK AML. Cyclin D3 has been previously reported to be involved in translocations in mature B-cell malignancies resulting in high expression levels of the gene[152]. Furthermore, increased expression of *CCND3* is also observed in HSCs transduced with *FLT3*-ITDs[178]. The patient with mutation in *CCND3* did not contain a *FLT3*-ITD.

Two patients carried a non-synonymous E253D substitution of *CCND3*

that was also present in their remission samples. The availability of corresponding remission blood or buccal tissue DNA is therefore a valuable tool for excluding rare polymorphisms as part of investigations into novel mutations; this will become an essential part of mutational studies if genes with very few mutations, typically substitutions are to be validated and included in molecular risk assessment.

FES

No mutations were detected within the tyrosine kinase domain of *FES*. This cannot, however, completely exclude the involvement of *FES* in leukaemia since mutation screening for this study was restricted to exons 13-19, the sites of previously reported mutations in colorectal cancer. However, mutations outside of this region may still be able to have an effect. Mutations or deletion within the first coiled-coil domain in the N-terminal region of *FES* activate the tyrosine activity. Additional mutations within the second coiled-coil domain result in loss of this activity[179]. These mutations are able to affect downstream *STAT3* signaling pathways. Therefore it is possible that mutations occur within these regions that were not detected by the screening process.

The investigation described herein examines the relationships between mutations in NK AML. Mutation screening of leukaemic blasts can yield valuable information for prognosis and therapeutic options. However, it does not yield sufficient insight into the sequential steps by which leukaemogenesis

occurs. It was therefore important to study the temporal process by which genetic damage accumulates during the development of AML.

CHAPTER 4: LATENCY AND ORDER OF MUTATIONAL EVENTS IN LEUKAEMOGENESIS

BRIEF INTRODUCTION AND DESCRIPTION OF STUDY

The increasing incidence of cancer with age suggests a gradual accumulation of genetic damage overtime for the disease to develop (**Figure 1.3**). In familial cases of AML, latency periods can last up to forty years or more before the disease is manifested. Current diagnostic techniques focus on the bulk of the disease, the leukaemic blasts. These blasts are the end point of leukaemogenesis and reveal little of the transformation process by which the disease arose. In this chapter we set out to examine the process of leukaemogenesis by studying the order by which mutations are acquired and the stage in haematopoiesis in which mutations occur. This was accomplished in two ways.

1. Examining the relationship between regions of acquired homozygosity in AML samples and mutations.
2. Using quantitative approaches to examine the existence and frequency of mutations in leukaemia stem cells and early progenitors.

ORDER OF GENETIC EVENTS IN LEUKAEMOGENESIS: ACQUIRED UNIPARENTAL DISOMY (aUPD)

DISCOVERY OF aUPD IN AML

Traditionally, G-banding has been the backbone of genetic examinations in leukaemia. However, more sensitive techniques have been developed which allow us to examine the disease more precisely at the molecular level.

The development of single nucleotide polymorphism (SNP) genotyping technology allows for the easy visualisation of loss of heterozygosity within the genome and for the detection of deletions or amplification of genomic regions (copy number)[180]. A SNP is a DNA sequence variation occurring when a single nucleotide differs between members of a species or between paired chromosomes in an individual and occur about once in every 1000 bases of the 3 billion bases in the human genome[181]. SNPs exist within coding sequences of genes, noncoding regions of genes, or in the intergenic regions between genes. SNPs in which both variants lead to the same polypeptide sequence are termed synonymous or silent while those which encode a different polypeptide sequence, are termed non-synonymous. The majority of SNPs (approximately 95%) either fall into the non-coding region of the genome or are silent, therefore do not have direct impact on protein sequence. However, SNPs that occur outside protein coding regions may still have consequences for gene splicing, transcription factor binding, or the sequence of non-coding

RNA and silent SNPs within the protein coding region may influence the kinetics of protein translation[182].

Gains and losses of chromosome number have long been observed in AML by karyotype analysis. More recently a novel abnormality in AML was revealed through the application of SNP genotype array technology[183]. The study by Raghavan *et al* revealed an acquisition of homozygosity without change in copy number. This aberration, which cannot be detected by conventional cytogenetic methods and was observed in approximately 20% of the AMLs studied.

This form of genetic lesion was termed *acquired uniparental disomy* (aUPD). Inherited uniparental disomy occurs when an individual receives two copies of a chromosome from a single parent with no copies from the other, resulting in homozygosity of the respective chromosome or chromosome region. aUPD is a somatic gain of homozygosity without gain or loss of chromosome material. The frequency by which this genetic lesion occurs in AML indicates that it may play an important role, or act as a signature in leukaemogenesis, even though it does not directly cause mutations in genes.

In AML, SNP profiling has detected several regions of aUPD, evident as large stretches of somatically acquired homozygosity. These regions usually continue from a variable location on the chromosome and continue to the telomere. The telomeric location of aUPD suggests a mechanism of mitotic recombination, because there is no net change in copy number in the affected

region (**Figure 4.1** and **4.2**). Mitotic recombination results in the generation of two daughter cells homozygous for distinct chromosomal regions which were subject to recombination. The homozygosity observed in AML is due to the outgrowth of one of the daughter cells likely due to a selective advantage conferred to the cell by a mutation within the region.

In studies carried out by Raghavan *et al,* sixty-four primary AML samples covering a cross section of FAB types and cytogenetic groups were analysed using 10K Affymetrix chips. aUPD was observed in 13 samples (approximately 20%).

Although the regions of homozygosity covered large portions of the chromosomes, we noted that within these regions are genes known to be mutated in AML. **Figure 4.3** shows an example of three cases of aUPD on 11p in which the gene *WT1* lies within the area of overlap. To investigate whether there was a potential relationship between gene mutation and somatically acquired homozygosity in AML, mutation screening was conducted on genes known to be mutated in AML located within the detected regions of aUPD.

HOMOZYGOUS MUTATIONS ARE FOUND IN REGIONS OF AUPD

For the thirteen patients with aUPD, mutation screening was undertaken for genes that lay within the region of homozygosity (**Table 4.1**). Due to the small number of samples, mutation screening was performed by amplification of the gene by PCR, followed by direct sequence analysis.

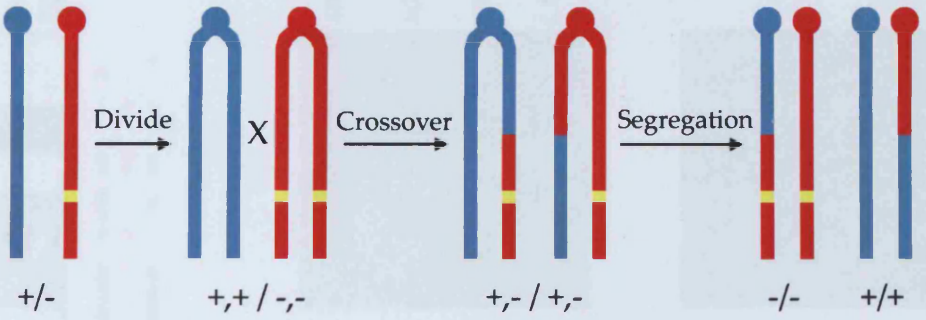


Figure 4.1: Loss of heterozygosity by mitotic recombination resulting in homozygous mutation

Cross-over between non-sister homologous chromatids followed by segregation results in homozygous mutants or wild-type cells. Alleles are represented in blue and red. A mutation occurs in one allele (yellow). When cell division takes place crossover of material occurs between non-sister homologous chromatids. When segregation of the alleles takes place, one daughter cell contains two copies of the mutant allele.

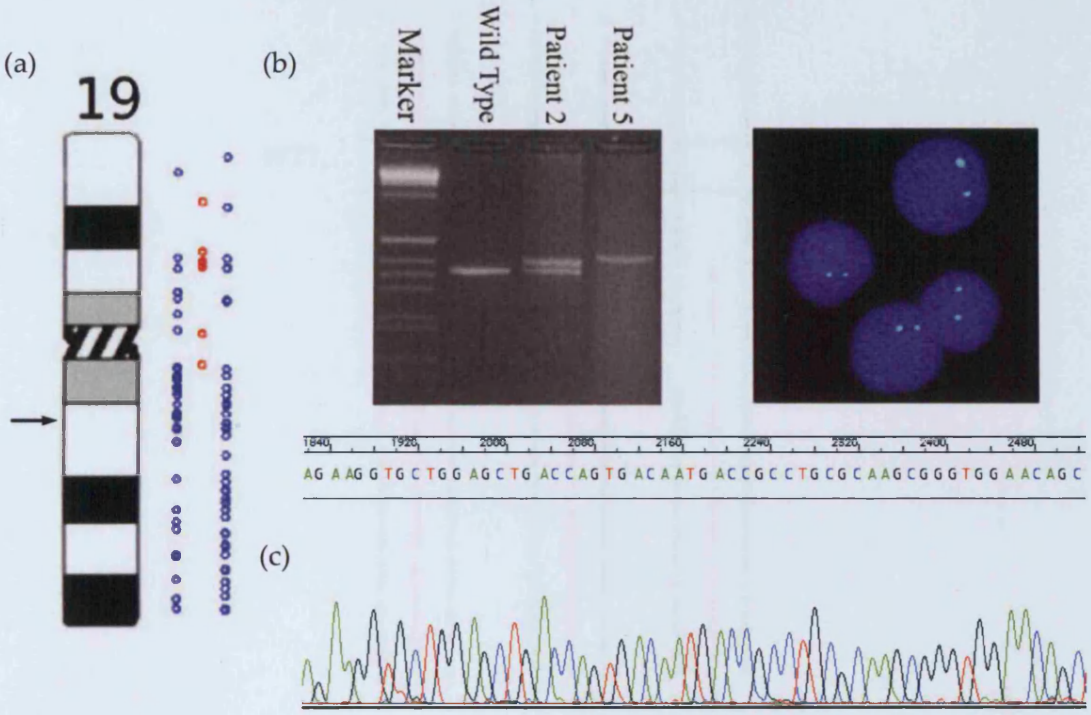


Figure 4.2: *CEBPA* mutation in association with aUPD 19q (taken from Snaddon et al)[118]

(a) Schematic of chromosome 19, blue dots indicates the presence of homozygous SNPs and red dots indicate the occurrence of heterozygous SNPs. Leukaemic blast DNA (patient 5) retains a large stretch of homozygousity covering the q arm of the chromosome which includes the *CEBPA* gene (location indicated by arrow).

(b) PCR gel analysis showing wildtype, heterozygous (patient 2) and homozygous (patient 5) gene mutation. FISH analysis using BAC clones show two copies of the CEBPA locus in patient 5, indicating that the homozygous 57bp ITD does not arise as a consequence of deletion of the wildtype allele.

(c) Sequencing trace of the homozygous mutation.

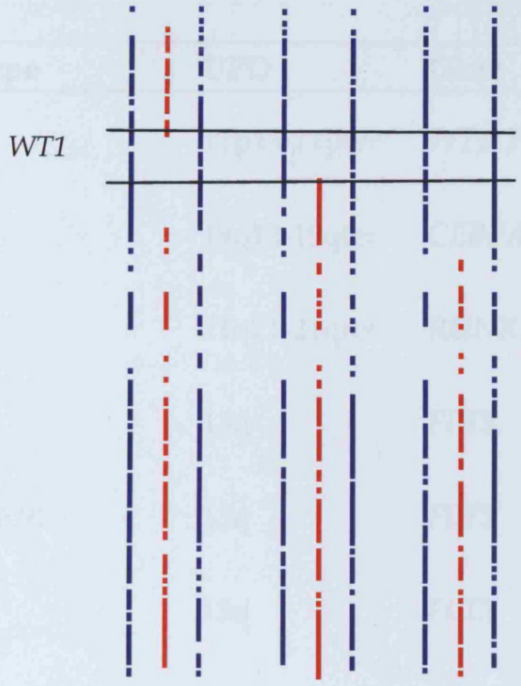


Figure 4.3: Regions of aUPD do not appear to be random events

Although the regions of homozygosity covered large portions of the chromosomes, they coincided with locations of genes known to be mutated in AML. The *WT1* gene is frequently mutated in Wilms' tumour and lies within the 11p region of overlapping aUPD. Only 1 of 3 of the above cases of aUPD on 11p showed mutation in *WT1* suggesting other genes or mutational mechanisms may be involved.

Patient	Karyotype	UPD	Gene	Mutation
1	46,XY	11p13-11pter	*WT1, PU.1*	1584 ins16bp *(WT1)*
2	46,XY	19q12-19qter	*CEBPA*	1038 ins57bp *(CEBPA)*
3	46,XY	21q21-21qter	*RUNX1*	970 insA *(RUNX1)*
4	46,XX	13q	*FLT3*	ITD 21bp *(FLT3)*
5	46,XY t(6;9)	13q	*FLT3*	ITD 63bp *(FLT3)*
6	46,XX	13q	*FLT3*	ITD 36bp/ 39bp* *(FLT3)*
7	46,XX	13q	*FLT3*	ITD 108bp *(FLT3)*
8	46,XX	11p11-11pter	*WT1, PU.1*	None
9	46,XX	11p11-11p14	*WT1, PU.1*	None
10	46,XX,dic(7;22) (q11.2;q10),+8	11q12-11qter	*MLL*	None
11	48,XY,+3,+10	6p21-6pter	*CCND3*	None
12	46,XX	6p11-6pter	*CCND3*	None
13	46,XY,der(12)t(1;12) (q11;p11.2)/46,XY	9pcen-9pter	*CDKN2A, CDKN2B*	None

Table 4.1: Patients, area of homozygosity, genes screened and result

aUPD was observed in 13 patients out of 64 tested. Genes known to be mutated in AML which are located within the areas of homozygosity were screened for mutations. Mutations were observed in 4 distinct loci, *WT1, CEBPA, RUNXI,* and *FLT3.* Numbering of mutations is according to the following Genbank accession numbers: NM_024426 (*WT1*), Y11525 (*CEBPA*), and NM_001754 (*RUNX1*).

Mutations were detected at four distinct loci (*WT1*, *FLT3*, *CEBPA* and *RUNXI*) in seven of the thirteen leukaemias examined (**Table 4.1**). In these seven cases, sequence analysis indicated a homozygous mutation. (**Figure 4.3**). Small background peaks of wild type sequence detected likely indicate infiltration of the tumour sample with normal cells. Four cases with aUPD across 13q were found to harbour homozygous internal tandem duplications in *FLT3*. The ITDs ranged from 21 to 108bp in length, with one sample demonstrating additional allelic evolution, with the additional insertion of 3bp. One case of 11p aUPDs was found to harbour a 16bp ITD in *WT1* (**Figure 4.4**). A single case of aUPD was observed on 19q, in which a 57bp ITD was observed in *CEBPA* (**Figure 4.2**). A homozygous 1bp insert was observed in *RUNX1* within an area of aUPD in 21q (**Figure 4.4**).

These results allowed us to develop a proposed timeline by which the LSC may have developed mutations. An initial "hit" occurs involving the gene-specific mutation on one allele. This is followed at a later time by mitotic recombination resulting in the mutation being rendered homozygous in a subset of cells. This homozygous mutation confers a selective advantage to the cell and it proliferates, giving rise to overt leukaemia.

In the case of *FLT3* internal tandem duplication mutations, the heterozygous state is a recognised poor prognostic risk factor for AML. Additionally, the loss of the wild-type *FLT3* allele occurs in 10% of patients with normal karyotype AML and is associated with a worse outcome[71]. This provides evidence that loss of heterozygosity can be viewed as a form of

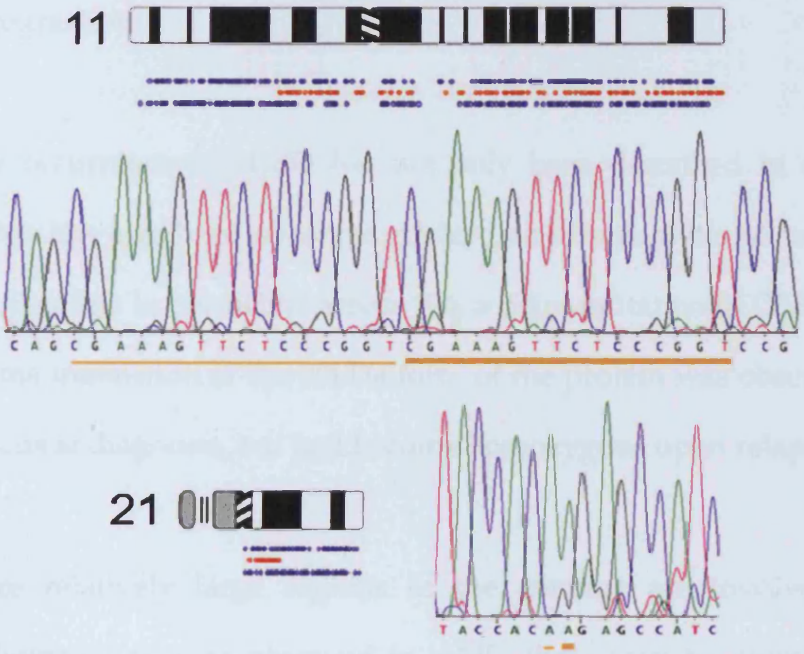


Figure 4.4: Sequencing traces of homozygous mutations discovered within areas of aUPD

Examples of homozygous mutations found within areas of acquired UPD. A homozygous 16 bp ITD in exon 9 of *WT1* was observed in a patient with aUPD across chromosome 11p. No mutation was detected in the other two cases. A single homozygous bp insertion in exon 5 of *RUNX1* was observed in a patient with aUPD on chromosome 21.

disease progression.

The occurrence of aUPD has not only been described in diagnostic samples, but has also been observed under pairwise analysis of relapse and remission samples. In one aUPD across 19q, a point mutation in *CEBPA* which results in the translation of the 30kDa form of the protein was observed to be heterozygous at diagnosis, but had become homozygous upon relapse.

Since relatively large regions of the genome are involved in the acquired homozygosity as observed in AML, there may be more than one potential mutational target in a given region. Investigation of 3 samples with aUPD of 11p revealed a mutation of *WT1* in only one case (**Figure 4.3** and **Figure 4.4**). Additionally, no mutations were observed in *PU.1*, which also located in this region, suggesting that other gene mutations may be implicated in the remaining two cases. Similarly, no mutations were detected in the candidate genes, *CCND3*, *CDKN2A* and *2B*, and *MLL* for aUPD on 6p, 9p, and 11q respectively, indicating there may be other targets contained within these regions.

TIMING OF MUTATION EVENTS IN LEUKAEMOGENESIS

DEVELOPMENT OF QUANTITATIVE REAL-TIME ASSAY FOR MUTATIONS IN *CEBPA*

With multiple mutational events occurring in normal karyotype AML, the order by which they arise may be important. The simplest way of studying

order of mutation is to examine whether there is concordance in mutation status between diagnosis and relapse samples. For example, mutations in *FLT3* are often not concurrent between presentation and relapse[184] indicating that mutations in this gene most likely represent secondary events. The same mutations in *NPM1* and *CEBPA* are maintained between presentation and relapse[185, 186] indicating that mutations in these genes are probably early initiating events.

In our centre, the availability of paired diagnostic and relapse material is limiting and so we have followed a different approach. AML is a stem cell disease which follows a similar hierarchy to normal haematopoiesis with the bulk of the disease driven by a LSC. Therefore, another way of examining order of mutational events is by studying in which HSC/progenitor compartment they occur. To do this, leukaemic cells were sorted into populations equivalent to stem cells and early progenitors in non-malignant haematopoiesis. These sorted cells were then tested for presence or absence of mutation.

To sort the leukaemic cells, the development of flow cytometery and the characterisation of cell surface markers have allowed the specific selection and study of different cell populations, including the LSC[187]. Functional approaches to studying the LSC have involved animal models. The murine xenotransplant NOD/SCID model has been an especially valuable resource in this respect, with the mice being able to faithfully recapitulate human disease[35]. The development of xenotransplantation technology has allowed for

accurate modelling of human AML growing within an animal system and the ability to characterise distinctive subpopulations of leukaemic cells.

Sensitive assays developed for specific mutations in AML, coupled with FACS technology to sort these rare leukaemic cells which have the ability to self renew and propagate the disease, can be used to determine whether a specific mutation exists within the earliest most potent cells. In this respect, investigation was undertaken to determine whether mutations in *CEBPA* are present within the leukaemic stem cell in normal karyotype AML. To achieve this, sensitive real-time quantitative mutation specific PCRs were developed and utilised for testing sorted leukaemic stem cells and early progenitors from patients for which the mutations were present in the blast compartment.

DESIGNING MUTATION SPECIFIC PCR

Patients with normal karyotype AML FAB type M1 were selected for study based on availability of diagnostic and/or relapse BM samples and on the presence of *CEBPA* mutations identified within during previous screening[118] (**Table 4.2**).

To develop an assay for *CEBPA* mutations, mutation specific forward primers were designed using the unique tumour specific sequences generated by the mutations. Insertion mutations introduce unique sequence arrangements at the junction of the insertion (**Figure 4.5**). Primers with the 3′ end lying across these regions were designed and tested for specificity. An

Patient	Age	FAB type	*CEBPA* mutation
1	37.9	M1	1138 ins (57bp)
2	38.2	M1	1081 ins (AGA)
3	19	M1	1091 ins (GCA)
4a	72.5	M1	363ins (GGCC)
4b			1096 ins (27bp)

Table 4.2: Characteristics of patients with *CEBPA* mutations included for real-time PCR analysis

Mutation specific real-time PCR assays were developed for four patients with mutations in *CEBPA*. Three of the patients had assays designed against mutations within the C-terminal region of the gene (Patients 1-3) The remaining patient contained biallelic N and C-terminal mutation for which a specific assay was developed for both. (Genbank Y11525)

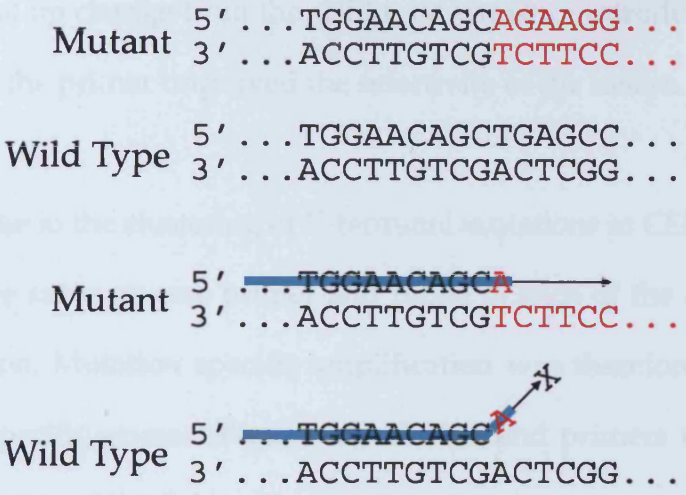


Figure 4.5: Strategy for design of specific primers for mutations in *CEBPA*

The duplication of 3-57bp in *CEBPA* results in the generation of a unique tumour specific sequence. Insertion mutation is shown in red, wild type sequence in black. Primers (blue bars) were designed to lie across these unique regions with the 3′ end of the primer designed not to anneal with the wild type sequence, therefore no polymerisation would occur.

additional bp change from the wild type sequence introduced 1-5 bp from the 3′ end of the primer improved the selectivity of the assays.

Due to the clustering of C-terminal mutations in *CEBPA*, it was possible to use the same reverse primer and probe in each of the assays designed for that region. Mutation specific amplification was therefore based on a single patient specific primer (**Figure 4.6**). Probes and primers were designed with the assistance of the PrimerExpress program.

This strategy was applied to mutations in 8 patients, including one patient with biallelic N and C-terminal region *CEBPA* mutations as well as the single base pair substitution contained within the Kasumi-6 cell line, a human acute myeloid leukaemia cell line established from an M2 patient. This cell line contains a point mutation in the C-terminal region of *CEBPA* which results in the altered protein structure of R305P.

Specificity of the assay was initially tested by amplifying wild type DNA and mutation containing DNA in a standard PCR reaction. If the primer set did not appear to amplify the wild type DNA, we proceeded to test the assay using quantitative real-time PCR (**Figure 4.7**). To help confer specificity, adjustments to primer length, identity of substituted base, and position of substituted base were made. In general 4 different base pair substitutions and 2 primer lengths were attempted for each mutation.

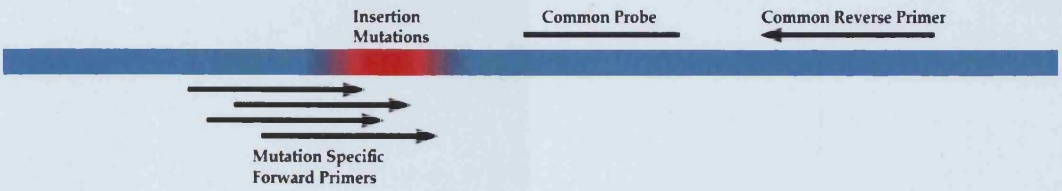
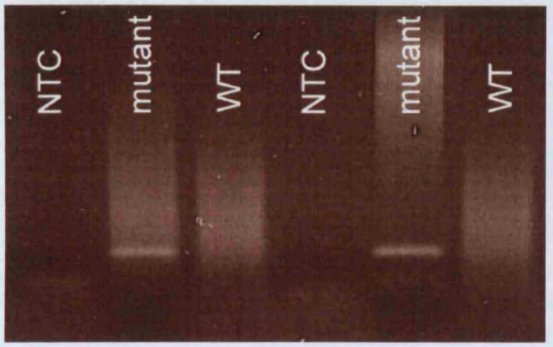


Figure 4.6: Schematic for development of mutation specific real time assays against mutations in the C-terminal region of *CEBPA*

Mutations in the C-terminal region cluster, allowing the use of a common probe and reverse primer. Mutation specific forward primers were developed, taking advantage of the unique tumour specific sequence generated by the corresponding insertion mutations.

a)



b)

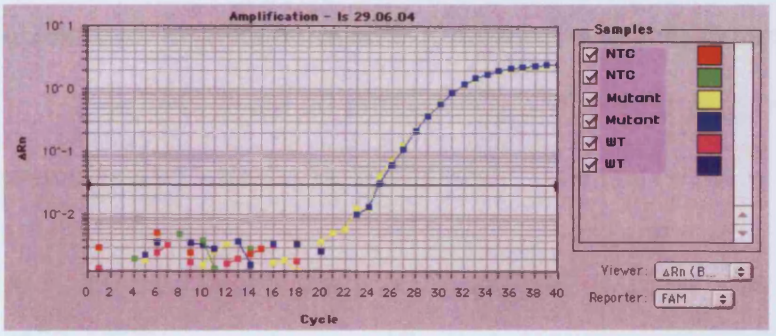


Figure 4.7: Testing of mutation specific assays for specificity

a) Real-time PCR with a master mix that does not degrade the final product was performed using the mutation specific primer set using DNA containing the mutation, wild type DNA and a non-template control (NTC). The product was visualised by gel electrophoresis. Only the sample which contained the mutation generated a product.

b) As in (a) the real time assay was conducted against DNA with and without the mutation. The mutation specific assays would not produce signal for WT DNA. Amplification was only observed in the duplicates shown in yellow and blue which contained mutant DNA.

SUCCESSFUL ASSAYS AND PATIENT DETAILS

Ultimately, five real-time quantitative PCR (RQ-PCR) assays were developed that selectively amplify mutated *CEBPA* DNA from 4 patients (**Figure 4.8**). Three patients (1, 2 and 3) had insertion mutations in the region of the gene that codes for the C-terminal DNA-binding/basic leucine zipper domain of the protein, the fourth contained mutations in both the N- and C-terminal regions of the C/EBPα protein (4a and 4b respectively). Mutations in 1, 2, 4a, and 4b are all tandem duplications ranging from 3 to 57bp in length while patient 3 has a 3bp insertion.

Designing a mutation specific probe

Another strategy that was attempted was to design a mutation specific probe. The probe was designed for a patient which had a 6 bp insertion of ACCCGC at nt 729 (Genbank Y11525). Since the ITD occurred in an area of highly repetitive sequence a unique sequence arrangement at the junction of the insertion was not generated. Therefore it was attempted to take advantage of the additional length of the sequence generated by designing a probe that would anneal only to the mutated DNA. The sequence of the probe covered the region 5′-ACCCGCACCCGC<u>ACCCGCA</u>CCCGCCGCCCGCGCACCT-3′ with FAM conjugated to the 5′ end and TAMRA to the 3′ end, the ITD is underlined. However when tested, this did not allow specific differentiation between wild type and mutant DNA.

Figure 4.8: Diagram of successful mutation specific assays

RQ-PCR approach for the specific detection of *CEBPA* mutations. For the C-terminal and N-terminal mutations, all patient sequences are shown (P1, P2, P3, P4b). Forward and reverse primer sequences are highlighted in yellow and purple respectively and Taqman probe location is indicated in green. A patient specific forward primer was designed to preferentially anneal to unique sequence generated by insertion mutation (red text and underlined) (see also Figures 4.4 and 4.5 for overall strategy). Base changes (lower case and blue highlighted) were introduced to the primer to prevent annealing and extension from WT sequence. The mean amplicon length was 158bp (range 150-164) for patient assays 1,2,3 and 4b. Amplicon length was 78bp for patient assay 4a.

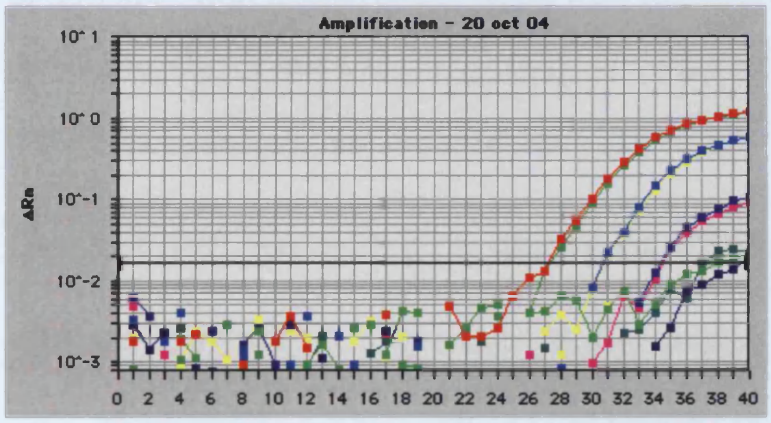
TESTING SPECIFICITY AND SENSITIVITY OF REAL-TIME ASSAYS

Real-time PCR analysis was performed on 25μl reaction mixture in duplicate using Taqman chemistry on the ABI PRISM 7700 Sequence Detector with default settings. Beta-2-microglobulin was used as the endogenous reference gene to assess total cell number per well.

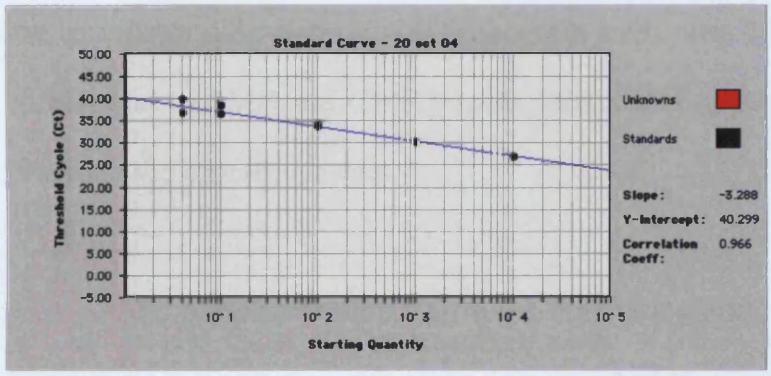
The specificity of each RQ-PCR assay was tested by amplifying 750ng (equivalent to 125000 cells) of each patient's diagnostic BM DNA, wt DNA, patient diagnostic BM DNA serially diluted in wt DNA and no template controls, in duplicate. An assay was deemed to be specific if no amplification was observed with wt DNA (**Figure 4.7**).

To test the sensitivity of each assay, a standard curve was generated by five serial dilutions of diagnostic DNA in water into 5 set standards (ranging from 125000 to 4 cells per well). The assays were also tested by diluting DNA containing mutation in wild type DNA and overall sensitivity was not affected (**Figure 4.9**). In all amplification plots, the same threshold was maintained, allowing comparison between experiments. The correlation coefficient of the standard curves was at least 0.95, indicating the assays had a high probability of precise quantification. The mean Ct for DNA equivalent to 125000 cells was 24.6 (range 24-27). The mean slope of the dilution curves was -3.2 (range -3.5 to -2.8) indicating good amplification efficiencies (the perfect theoretical slope is –3.3). The maximum reproducible sensitivity (defined as the dilution for which duplicates were both positive) for all the patient mutation assays was

a)



b)



c)

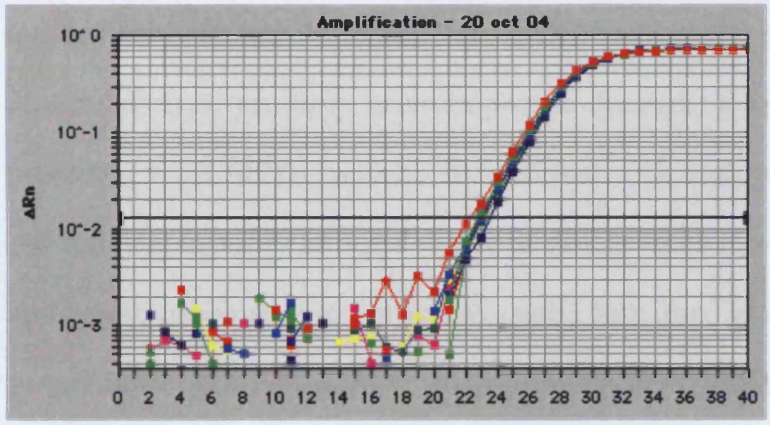


Figure 4.9: Testing sensitivity and specificity of real time assays. Example of standard curves for P2

a-b) Serial dilution of patient DNA in wild type DNA showing *CEBPA* mutation specific assays.

c) Same assay with β-2 microglobulin control to measure total DNA content.

10^{-4} (8.0×10^{-5}; 10 in 125000) and therefore were of sufficient sensitivity for MRD monitoring $<10^{-4}$ (**Table 4.4**).

DETERMINING THE POINT OF ORIGIN OF *CEBPA* MUTATIONS, THE LEUKAEMIC STEM CELL AND EARLY PROGENITORS

These assays were also utilised to address the question of where within the primitive populations of leukaemia cells, the LSC and subsets that correspond to CMP and GMP populations in non-malignant haematopoiesis, in which the mutation arises.

AML peripheral blood samples from 4 patients were flow sorted into three distinct hierarchal compartments. (i) The LSCs, which correspond to the most primitive haematopoietic stem cells that have the potential for long-term repopulation of multiple haematopoietic lineages; (ii) the common myeloid progenitors (CMPs) which in normal haematopoiesis have limited self-renewal capacity and the ability to differentiate into all myeloid cell types and their downstream progeny, and (iii) the granulocyte/macrophage progenitors (GMPs), which in non-malignant cells, also have limited self-renewal capacity and restricted in their differentiation ability to granulocytes and macrophages.

These populations were sorted based on the lack of mature lineage markers (CD2, CD3, CD14, CD16, CD19, CD24, CD41, CD56, CD66b and

Assay	Mutation	Sensitivity	Ct 125k	Ct 10	Slope	correlation
1	1138 ins (57bp)	8.0 e-5	25	38	-3.4	0.994
2	1081 ins (AGA)	8.0 e-5	26	37	-2.8	0.945
3	1091 ins (GCA)	8.0 e-5	24	37	-3.2	0.991
4a	363ins (GGCC)	8.0 e-5	24	37	-3.2	0.986
4b	1096 ins (27bp)	8.0 e-5	25	37	-3.0	0.981

Table 4.4: Summary of sensitivity and specificity testing of real-time PCR analysis for *CEBPA* mutations

A standard curve was generated by five serial dilutions of diagnostic DNA in water (ranging from 125000 to 4 cells per well). In all amplification plots, the same threshold was maintained, allowing comparison between experiments. The correlation coefficient of the standard curves was at least 0.95, indicating that the assays had a high probability of precise quantification. The mean Ct for DNA equivalent to 125000 cells was 24.6 (range 24-27). The mean slope of the dilution curves was -3.2 (range -3.5 to -2.8) indicating good amplification efficiencies (the perfect theoretical slope is –3.3).

glycophorin A). In addition, the majority of LSCs express CD34 cell surface marker, but not CD38 (Lin-/CD34+/CD38-). The CMPs and GMPs express CD34, CD38, and CD123 and are distinguished by presence (GMP) or absence (CMP) of CD45RA (as described in Chapter 2)[187]. Unfortunately due to limited availability of patient samples, there was insufficient material for re-analysis of the populations by FACS before isolating DNA for mutation analysis.

DNA isolated from each compartment was tested with the RQ-PCR assays for mutation. In the stem cell compartment (Lin-/CD34+/CD38-), *CEBPA* mutation was detected in all patients, with the exception of patient 1 and the second mutation in patient 4 where insufficient cells were collected for testing. The downstream populations, CMPs and GMPs, tested positive for the mutations in all cases, except for patient 3 where insufficient cells were collected for the CMP population. This data is summarised in **Table 4.5**.

These assays show that the *CEBPA* mutations are present within the stem cell compartment in evaluated patients; however the ability to accurately quantify the amount of mutation present in these compartments is limited. Due to limited amounts of sample available for sorting and testing, the sorted populations were not reanalysised for purity. Therefore it is possible for the tested populations to contain cells that do not strictly fall within the target population. Furthermore, the BM tested had an average blast count of 70.25% (range 40-95%) therefore the amount of mutation could also be diluted by presence of non-leukaemic cells.

		HSC	CMP	GMP
P1	n=2	n/a	11%*	21%
P2	n=1	34%	100%	100%
P3	n=2	17%*	n/a	40%
P4a	n=2	100%*	16%	53%
P4b	n=2	n/a	15%	54%

n/a: Insufficient amount of sample was recovered for testing

* Mutation testing was only successful for one of the two cases

Table 4.5: Mutation detection in sorted AML samples

AML peripheral blood samples from 4 patients were flow sorted into three distinct hierarchal compartments, equivalent to the HSC, CMP and GMP in normal haematopoiesis. DNA isolated from each compartment was tested with the RQ-PCR assays for mutation using β2M as a control for numbers of cells tested. The average percentage of cells from the sorted populations is indicated for each patient.

ENGRAFTMENT OF AML SAMPLES

A further approach to determine whether mutations within *CEBPA* occur within a primitive cell type was to utilise the *in vivo* NOD/SCID model of leukaemia engraftment. As discussed in the introduction to this thesis, the ability of human cells to engraft within the murine host indicates that cells injected have the ability to self-renew. With AML, when engraftment occurs, the disease is faithfully reproduced, giving rise to cells with the same phenotype as the original disease.

Leukaemia samples from patients with RQ-PCR assays for *CEBPA* mutation were sorted into the stem cell compartments described above and injected into NOD/SCID mice to examine the population's engraftment potential. Following engraftment, the cells would be recovered and then be tested for presence or absence of mutation. This would demonstrate if the mutation is contained within a primitive cell type which is able to repopulate the mouse bone marrow.

AML cells from P4 were sorted by flow cytometry into Lin-/CD34+/CD38-, CMPs, and GMPs populations. The recovered cells were then injected into sub-lethally irradiated NOD/SCID β2 mice. The mice were maintained under normal care conditions with the addition of antibiotics for 12 weeks. The mice were then killed, their bone marrow collected and analysed by flow cytometry for engraftment of human cells (**Table 4.6**). Due to an infection of the mouse colony several mice died before the experiment had

Fraction	Mice injected	Mice Survived	Engraftment
Lin- 34+ 38-	4	1	0/1
GMP	5	2	0/2
CMP	4	2	0/2
Whole Mononuclears	5	5	0/5

Table 4.6: LSC and early progenitors from P4 were sorted and injected into NOD/SCID β2M mice

Cells were sorted into HSC, CMP and GMP compartments and injected into lethally irradiated NOD/SCID β2M mice. Mice were left for 12 weeks and then bone marrow was harvested and examined for the engraftment of human cells.

run its course. No human engraftment was observed from any compartment (**Figure 4.10**).

To confirm that the patient had a non-engrafting phenotype, whole mononuclear cells from the patient were injected into 5 NOD/SCID β2 mice, with 100,000 cells per mouse. After 12 weeks the mice were killed and analysed. Once again no engraftment was observed (**Figure 4.11**).

As has been described by our group during the duration of this thesis[47], the ability of a particular AML to engraft in the NOD/SCID model is related to the prognosis of individual AML cases. In general it was found that AMLs with favourable prognosis were unable to engraft within the murine model. Since mutations in *CEBPA* are associated with favourable prognosis within the normal karyotype subset of AML, our results are consistent with this observation.

SUMMARY

In contrast with Chapter 3 where we restricted analysis to the occurrence of mutation in the myeloblast Chapter 4 has addressed the questions as to the order by which these mutational events might arise. The data has provided fresh insight into how changes in gene dosage arises (throught a mechanism of aUPD), have reestablished the importance of *WT1* mutation as a cause of AML, shown the application of mutation specific PCR to the study of MRD and its parallel use to investigate the cellular

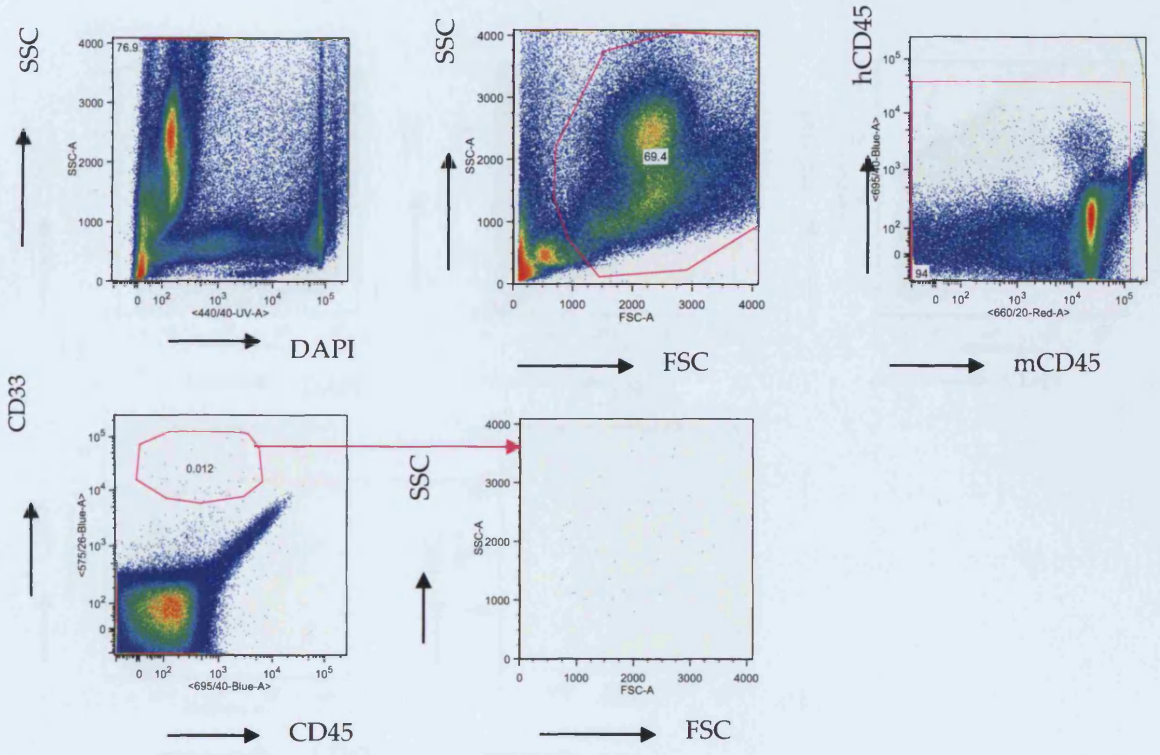


Figure 4.10: Sorted P4 cells did not engraft

Analysis of mouse bone marrow 12 weeks post injection of sorted P4 cells. Cells were examined for expression of human CD45 and CD33. The small population of events which was positive for both markers did not represent true engraftment.

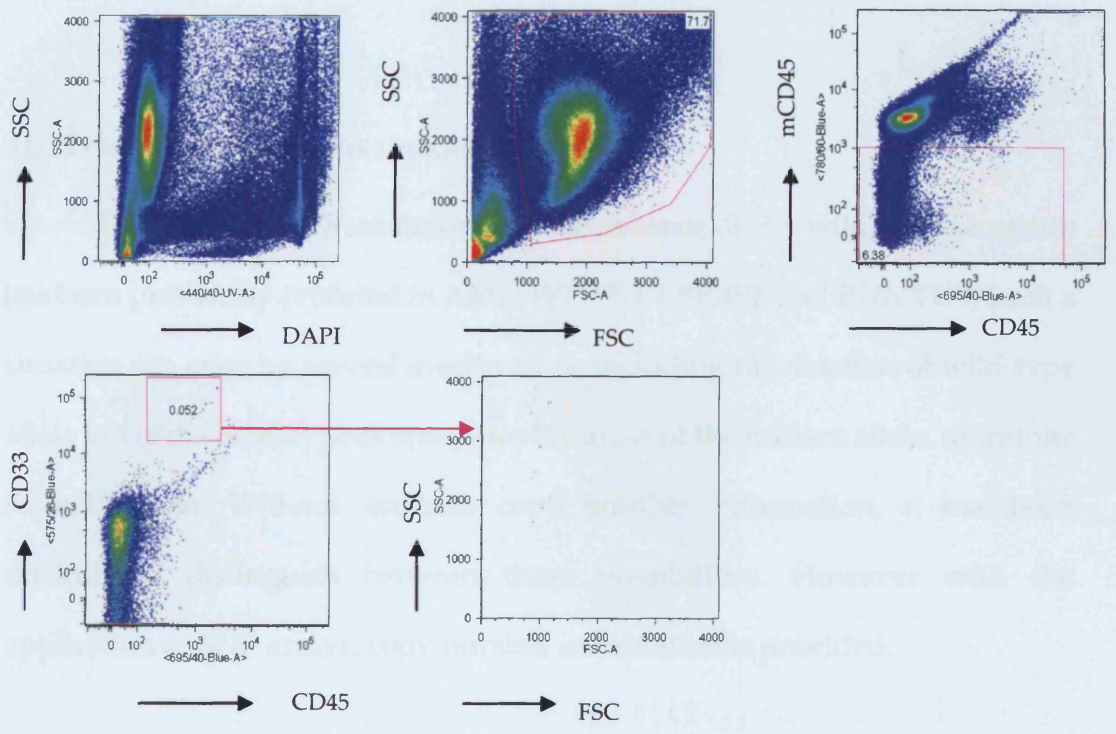


Figure 4.11: Whole mononuclear cells from P4 do not engraft

Analysis of mouse bone marrow 12 weeks post injection of whole mononuclear P4 cells. Cells were selected for lack of expression of murine CD45 and then for dual expression of human CD45 and CD33. The small population of events which was positive for both markers did not represent true engraftment.

compartment in which mutation arise. These new data have changed how the group views AML and has laid the foundation for the current direction of the groups' studies.

aUPD AS A MECHANISM OF LEUKAEMOGENESIS

The presence of a mutation and the absence of the wild-type sequence has been previously reported in AML: *WT1*[188], *CEBPA*[114], and *RUNX1*[189]. Such a situation can arise by several mechanisms, including the deletion of wild-type allele to yield a hemizygous state, amplification of the mutant allele, or mitotic recombination. Without accurate copy number information, it has been difficult to distinguish between these possibilities. However with the application of SNP arrays, copy number information is provided.

Not all stretches of aUPD in this study were found to contain mutations in the genes examined, indicating that there must be other factors that underlie the outgrowth of clones containing these homozygous regions. In that respect, the study of aUPD can provide clues into the location of novel genetic events which are important in leukaemogenesis.

In solid tumours, acquisition of homozygosity without loss in copy number has been observed in breast cancer[190, 191], uveal melanoma[191], Wilms' tumours[192], and retinoblastoma[193]. In addition, the background rate of mitotic recombination has been measured[194] in normal human and mouse cells to be approximately 10^{-4} to 10^{-5} Thus, given the widespread nature of mitotic

recombination, it is therefore very probable that the acquisition of homozygosity has a similar role in other tumours, besides leukaemia.

CHALLENGES AND APPLICABILITY OF DEVELOPING MUTATION SPECIFIC ASSAYS

Although *CEBPA* is a gene that plays an important role in AML there are difficulties involved in designing specific assays for mutations in this gene. Unlike genes such as *JAK2* or *NPM1*, where the mutations found within the gene are the same from case to case, mutations in *CEBPA* are variable and therefore require specific design for each. This difficulty is partially alleviated by the clustering nature of mutations in *CEBPA* and the prevalence of insertion mutations, for which the assay design is applicable.

As about 60% of *CEBPA* mutations in AML reported in the literature[77] are insertions or deletions this makes them an attractive RQ-PCR target. Indeed a survey of the sequences shows clustering of mutations and indicates that Probe 1, 2, 3, 4b described herein could be applied to about 40% of these. Many of these mutations result in the insertion or deletion of a small number of additional nucleotides that should allow for the design of a specific primer against the unique sequence generated by the mutation[67].

Unfortunately, lack of appropriate remission and relapse samples made it difficult to test the mutation specific PCRs for application as a MRD monitoring technique. However, with favourable prognosis it is of potential therapeutic significance to stratify patients containing *CEBPA* mutations and

monitor them over the course of their disease. Its application and correlation with clinical outcome may now be evaluated, alongside other approaches assay such as multiparametric flow cytometry and *WT1*-expression changes, for its ability to quantify MRD in AML informatively. Applying the same principles mutations in other genes also have potential as appropriate targets that may increase the number of AML patients with normal karyotype amenable to MRD monitoring.

CEBPA MUTATIONS IN STEM CELL AND EARLY PROGENITORS

As studies have shown that the location of mutations and genetic events can play an important role in AML development it is important to determine where these key events take place. In our study, leukaemia samples were sorted into LSC, CMP and GMP compartments and then tested for presence of *CEBPA* mutations by specific real time PCR analysis. In general, *CEBPA* mutations were detected in all compartments for all samples.

Unfortunately it was not possible to examine further whether mutations in *CEBPA* did indeed take place within long-term repopulating cell type due to the lack of engraftment from the AML samples tested. Our lab has recently reported that the ability of a particular AML to engraft in the NOD/SCID model is related to the prognosis of individual AML cases which will be discussed further in the final chapter.

CHAPTER 5: FUNCTIONAL CONSEQUENCES ON HAEMATOPOIESIS OF MUTANT FORMS OF *CEBPA*

BRIEF INTRODUCTION AND DESCRIPTION OF STUDY

Through the course of this thesis, we have examined the pattern and relationship between mutations as well as the timeline through which mutations give rise to AML. However, understanding the function of these mutations is an essential next step in addressing their role in leukaemogenesis.

Although *CEBPA* is a well studied transcription factor, the molecular mechanisms underlying the mutations of this gene and its role in leukaemogenesis have not been clearly defined. We have chosen to investigate this gene for several reasons. Mutations in *CEBPA* are the third most common genetic defect observed in normal karyotype AML, occurring in approximately 15% of NK AML patients[77, 114, 143]. *CEBPA* mutations have been implicated in familial AML[100-102], and remain consistent between diagnosis and relapse[119] indicating that these are likely to be early events in leukaemogenesis. Biallelic mutations in *CEBPA* are common and in our mutation studies we also noted that both mutations occurred on the same allele in some patients.

Familial cases of AML involving mutations in *CEBPA* provide particular insight into the importance of the two types of mutations observed

in this gene. Recently, our laboratory described a familial case of AML involving mutations in *CEBPA*[100]. The latency period for development of AML in this family ranged from 10 to 30 years. The family contained a heterozygous germline mutation of del (C) at nucleotide 212 (GenBank DNA sequence number Y11525; Swiss-Prot protein sequence number P49715). The mutation resulted in the premature termination of the C/EBPα protein at codon 158 and the selective production of the 30kDa isoform by way of an internal ATG site. In this chapter, these out of frame mutations which cause the selective translation of the 30kDa isoform will be referred to as N-terminal mutations.

In one affected family member, a second *CEBPA* mutation, an internal tandem duplication resulting in disruption of the DNA binding/dimerisation domain of the protein, was detected. This mutation was observed to have been somatically acquired in addition to the germline mutation. These in-frame mutations which disrupt the basic leucine zipper (DNA binding/dimerization) region of the protein will be referred to as C-terminal mutations. It is important to note that similar germline and somatic events have since been noted to occur in other familial pedigrees.

As in familial cases, biallelic mutations of *CEBPA* with an N-terminal mutation on one allele and a C-terminal mutation on the other allele are found in somatic AML[119, 195]. Furthermore, 2 mutations on the same allele have been reported by our laboratory and others[119] (**Figure 5.1**). In this part of the project we have examined the significance of the acquisition of two distinct mutations, N and C-terminal, on the same allele and their affect on normal

Figure 5.1: Biallelic mutations in *CEBPA* are observed in AML

Cloning the gene can reveal information about the nature of the mutations. Biallelic mutations with N and C terminal mutations on separate alleles have been observed as well as the mutations occurring on the same allele.

haematopoiesis. To investigate the transforming properties of *CEBPA* mutations and to test the significance of the double mutation variant of the gene, *CEBPA* containing an N-terminal mutation and both N and C-terminal mutations were cloned into lentivirus vectors and used to transduce lineage-depleted cord blood mononuclear cells. These cells contain the HSC and early progenitors of the haematopoietic system. Using transduced cells, *in vitro* colony forming cell assays and *in vivo* engraftment assays were performed to determine the effect of *CEBPA* mutations on normal haematopoietic proliferation and differentiation.

CLONING MUTANT FORMS OF *CEBPA*

A patient with biallelic *CEBPA* mutations was selected as a template for functional analysis of dual N and C-terminal mutations. The patient contained an insertion at base 363 of four base pairs GGCC (N-terminal mutation) and an in-frame internal tandem duplication of 27 base pairs at 1096 (C-terminal mutation) (GenBank Y11525). Real-time quantitative PCR assays for the mutations in this patient were previously developed (see Chapter 4, P4 for details).

As mentioned previously, familial cases involving *CEBPA* mutations begin with a germline N-terminal mutation and subsequently acquire a C-terminal mutation. To emulate this, two mutant constructs were cloned, one containing the N-terminal mutation ("N" construct) and one with both mutations ("NC" construct). In these studies, we did not attempt to eliminate

or downregulate normal *CEBPA* expression.

The cloning strategy for the creation of a lentivirus containing the mutant *CEBPA* gene is outlined in **Figure 5.2** and described in detail in chapter 2. Briefly, *CEBPA* from the patient was cloned into Topo vector and screened for mutations. The appropriate constructs were then cloned into an entry vector (pENTR1A) containing IRES-eGFP. The mutant gene and IRES-eGFP was next cloned into a lentivirus vector, resulting in the final construct of mutant *CEBPA* and GFP. The choice of a lentiviral vector reflects their ability to transduce non-proliferating cells, therefore this virus delivery system was chosen to target the primarily quiescent HSCs.

In addition to the two mutant constructs ("N" and "NC") and the mock (IRES-eGFP only), a wild type *CEBPA* and a construct containing only the C-terminal mutation were cloned. Despite several attempts, the presence of additional point mutations in the C-terminal mutation construct rendered this construct unusable for the remaining experiments. The generation of an appropriate C-terminal construct forms part of ongoing experiments in the laboratory.

TESTING MUTANT *CEBPA* LENTIVIRUS CONSTRUCTS FOR PROTEIN EXPRESSION

To ensure that the correct protein forms of C/EBPα was produced by the lentivirus vector, 293T cells were transiently transfected with the *CEBPA* constructs. Cells were harvested two days after transfection and protein

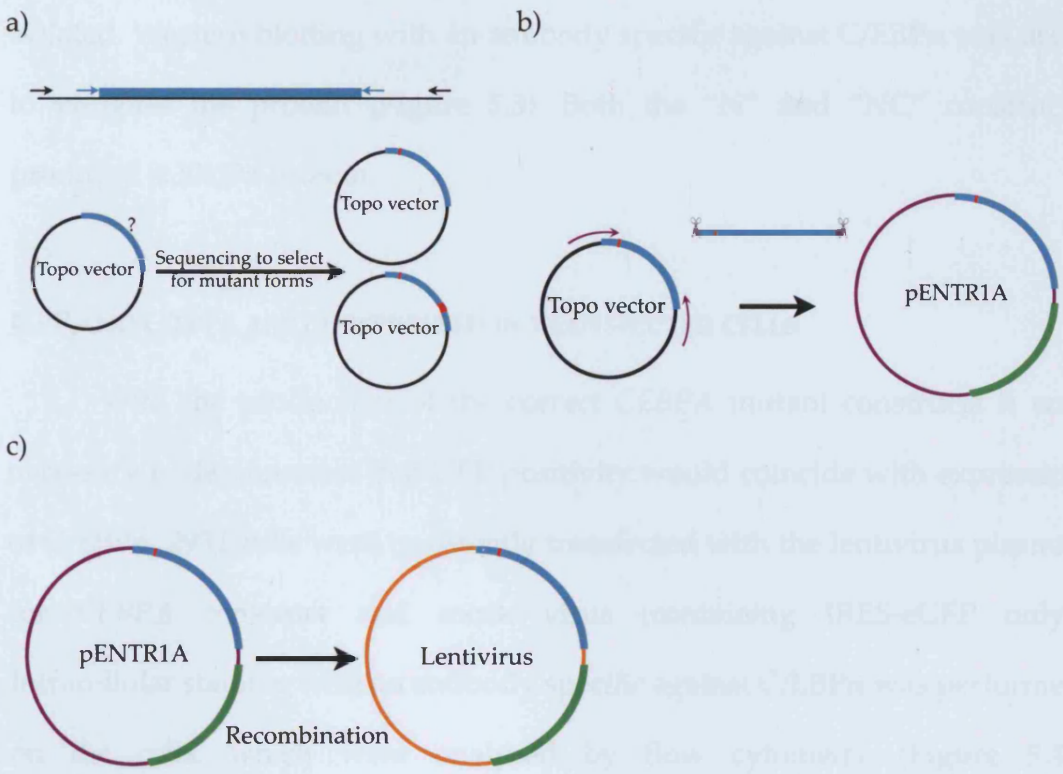


Figure 5.2: Cloning strategy for production of lentivirus containing *CEBPA* mutations

a) *CEBPA* was amplified directly from myeloblast genomic DNA by way of nested PCR approach, using external primers (black) and internal primers (blue). This PCR product was ligated into Topo 2.1 vector (Invitrogen) and transformed into Top10 bacteria. Colonies positive for successful ligation were selected, grown and plasmid DNA was extracted. The authenticity of the clones was confirmed by direct sequencing analysis.

b) PCR was performed using the Topo vectors as template to insert a FLAG sequence to the C-terminal portion of the protein and appropriate restriction enzyme sites to allow for digestion and ligation into a linerised pENTR1A vector which contains IRES-eGFP gene. The flag sequence and GFP were included to allow for visualisation of the gene during the cloning and transduction process.

c) The mutant *CEBPA* gene along with IRES-eGFP was next recombined into the lentivirus vector producing the final mutant construct.

isolated. Western blotting with an antibody specific against C/EBPα was used to visualise the protein (**Figure 5.3**). Both the "N" and "NC" constructs produced a 30kDa protein.

GFP AND C/EBPA ARE COEXPRESSED IN TRANSFECTED CELLS

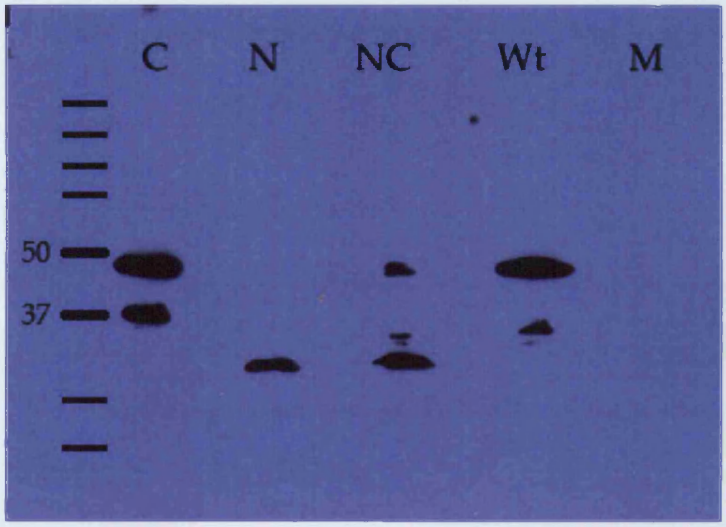
With the production of the correct *CEBPA* mutant constructs it was necessary to demonstrate that GFP positivity would coincide with expression of C/EBPα. 293T cells were transiently transfected with the lentivirus plasmid for *CEBPA* construct and mock virus (containing IRES-eGFP only). Intracellular staining with an antibody specific against C/EBPα was performed on the cells, which were analysed by flow cytometry (**Figure 5.4**). Approximately 20% of cells expressed GFP for both viruses. The majority of the GFP positive cells for the *CEBPA* construct were also positive for C/EBPα, demonstrating that cells which express GFP also express C/EBPα.

FUNCTIONAL CONSEQUENCES OF MUTANT *CEBPA* ON LINEAGE DEPLETED CELLS

The aim of this study was to determine the effect of *CEBPA* on normal haematopoiesis using mutations observed in patient samples, and to ascertain if the combination of mutations is important.

HSC studies rely on several standard *in vivo* and *in vitro* assays for the assessment of haematopoietic stem cell activity. Presently, the best model used

a)



b)

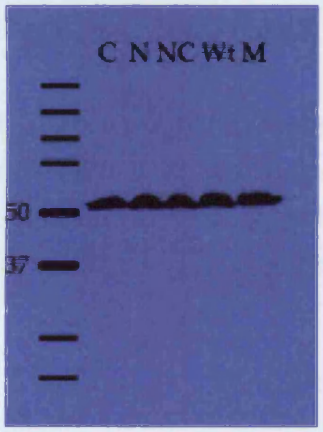


Figure 5.3: Western blotting to confirm expression of C/EBPα

The mutant lentivirus constructs were tested for C/EBPα expression by western blotting. 293T cells were transiently transfected with the mutant plasmids as described. Protein was isolated and run on a 12% SDS-PAGE gel with a Bio-Rad Kaleidoscope ladder for molecular weight determination and visualised by western blotting with a C/EBPα specific antibody (a) and β-tubulin for loading control (b).

Both "N" and "NC" constructs contain a frameshift mutation in the N-terminal region of the protein which forces the production of the 30kDa isoform. This isoform was observed for both constructs. The "Wt" construct, containing wild type *CEBPA* and the "C" construct, containing only the C-terminal region ITD, both produced the 42kDa isoform. Cells transduced with the mock "M" construct which contains only GFP does not express C/EBPα.

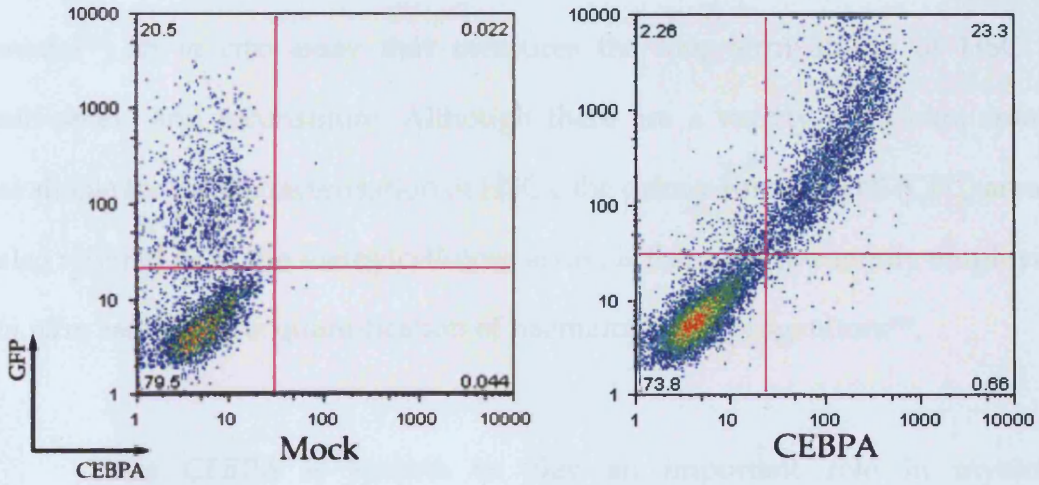


Figure 5.4: C/EBPα and GFP are co-expressed

293T cells transiently transfected with lentivirus plasmid expressing mutant *CEBPA* and IRES-eGFP show co-expression of C/EBPα and GFP. Cells were fixed and permeabilised following 48 hours of transfection, stained with an antibody specific for C/EBPα and analysed by flow cytometry.

to define the repopulating capacity of human HSCs is the NOD/SCID mouse model[196], an *in vivo* assay that measures the long-term ability of HSC to self-renew and reconstitute. Although there are a variety of *in vitro* assays available for the characterisation of HSCs, the colony-forming cell (CFC) assay, also referred to as the methylcellulose assay, is the most commonly employed *in vitro* assay for the quantification of haematopoietic progenitors[197].

Since *CEBPA* is known to play an important role in myeloid development in a variety of cell compartments (HSC, CMP, GMP), we decided to examine the effects of the mutant forms of *CEBPA* constructs on primitive lineage depleted cord blood mononuclears (Lin- MNCs), which contains all of the aforementioned subsets.

TRANSDUCTION OF LINEAGE DEPLETED CORD BLOOD CELLS

A schematic of the strategy used in these experiments is shown in **Figure 5.5.** Lin- MNCs were transduced with virus containing various mutant combinations of *CEBPA* (details of methodology are given in chapter 2). These cells were then applied to the NOD/SCID xenotransplant model to assess the effects of mutant *CEBPA* on long-term repopulation ability of the cells and to methylcellulose assays to assess short-term repopulation and differentiation effects. Remaining cells were tested for GFP expression by flow cytometry to determine transduction efficiency.

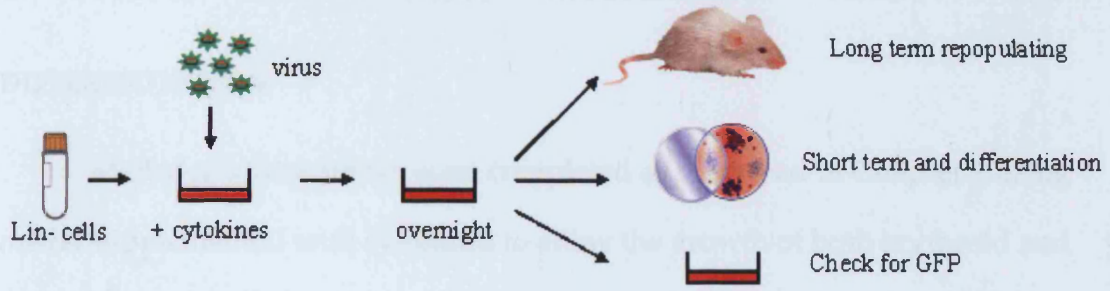


Figure 5.5: Schematic for set up of functional assays to test for the effects of mutant *CEBPA*

Lineage depleted cord blood mononuclear cells were incubated with cytokines and transduced with lentivirus containing the mutant forms of *CEBPA*. Cells and virus were incubated overnight (16h) and then the cells were used in the following assays. Long term repopulating capacity of the cells was tested by injecting NOD/SCID mice with cells transduced with virus expressing the mutant constructs. Short term repopulation and differentiation was tested by CFC assays. Transduction efficiency was tested by examining GFP expression of transduced cells.

IN VITRO: SHORT TERM REPOPULATING ABILITY AND DIFFERENTIATION

Methylcellulose assays were completed as described in Chapter 2 using media supplemented with cytokines to allow the growth of both erythroid and myeloid colonies. To reduce variation, Lin- MNCs from the same preparation were used within each assay. Non-transduced cells from the same aliquot of Lin- MNCs were used as a negative control.

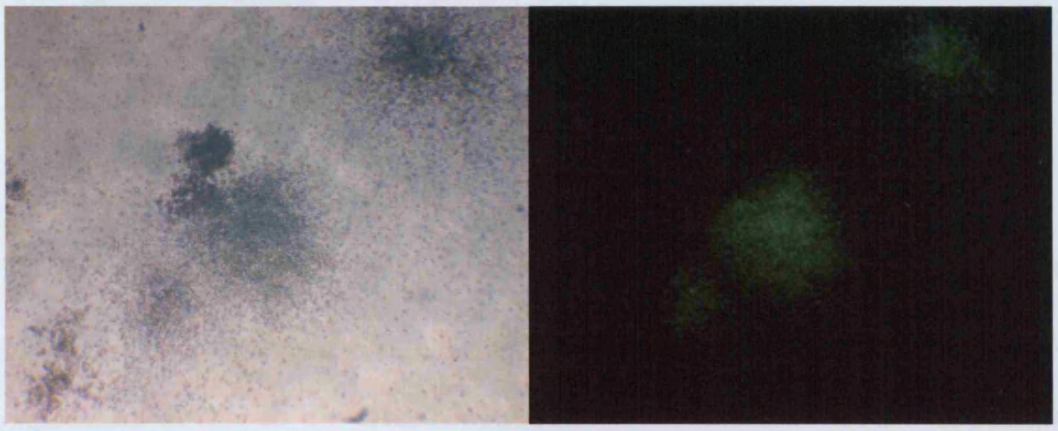
Colonies on each plate were scored based on morphology of the colonies into 5 categories; BFU-E, CFU-E, GM, G and M (as described in Chapter 2). Initially a total colony count was performed. GFP positive colonies were then also scored based on colony morphology.

ANALYSIS OF METHYLCELLULOSE ASSAYS

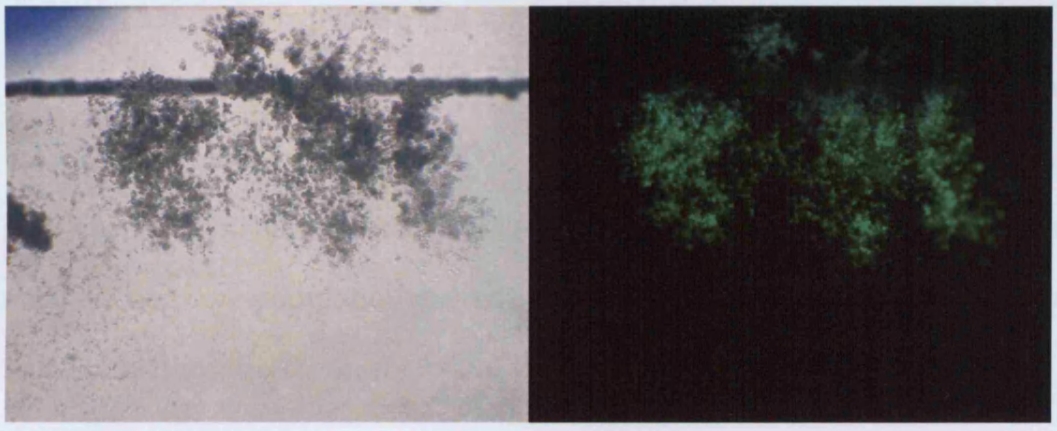
GFP visualisation

In total, four *in vitro* assays were carried out testing the "N" and "NC" constructs. Colonies were counted using a Lieca DM1L inverted microscope under 50x magnification. GFP was visualised by excitation from a UV lamp. **Figure 5.6** shows examples of GFP positive colonies from various constructs. While some GFP positive colonies are brighter than others, they are clearly distinguishable from non-GFP expressing constructs, as demonstrated in **5.6a** and **c** in which GFP negative colonies are observed in the same field of view as

a) GFP positive myeloid colony from N mutation construct



b) GFP positive erythroid colony from N and C mutation construct



c) GFP positive myeloid colony from mock construct

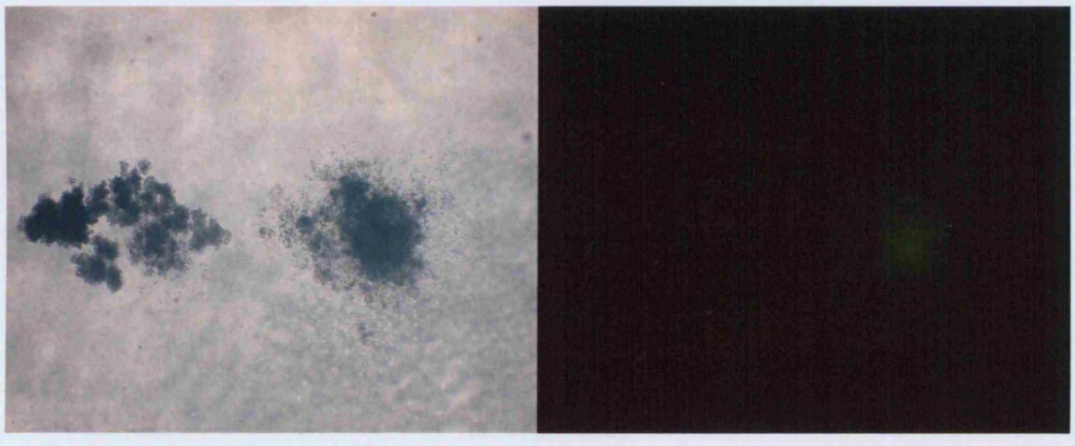


Figure 5.6: Examples of GFP expressing methylcellulose colonies

GFP expression was visualised using an inverted fluorescence microscope. While some GFP positive colonies are brighter than others, they are clearly distinguishable from non-GFP expressing constructs, as can be seen in a) and c) where GFP negative colonies co-exsist in the same field of view.

GFP positive colonies.

CHECKS AND CONTROLS

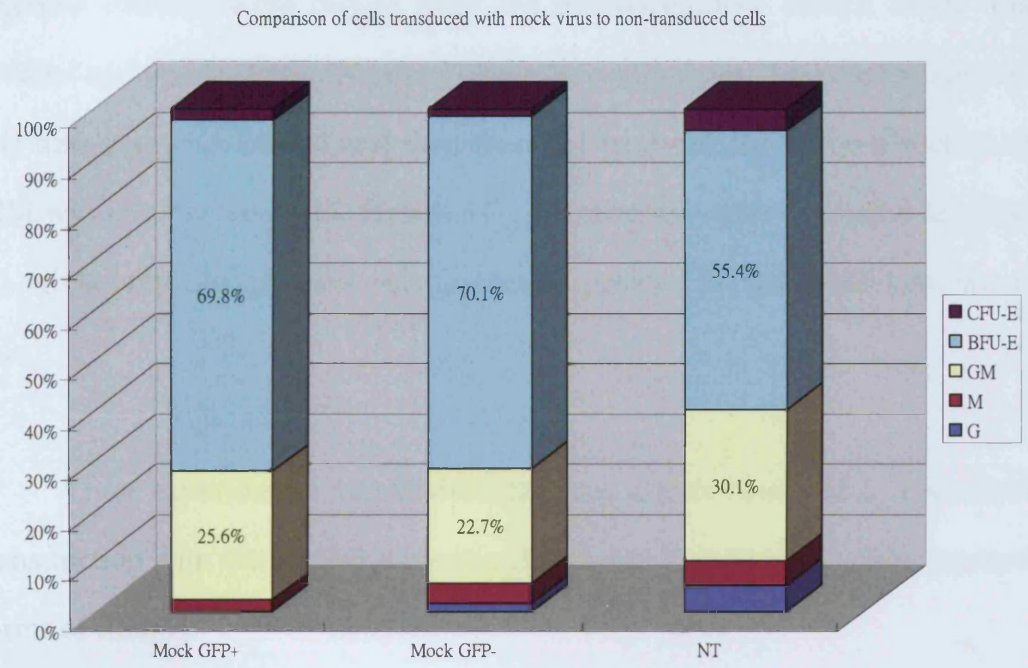
The lentivirus and GFP do not have an effect on proportion of colony types

To ensure that the effect observed within the assay was not due to GFP expression or due to the viral construct itself, transduction of the Lin- MNC cells was also performed using a viral construct containing only IRES-eGFP (mock). A total colony count of the plate was performed, assigning each of the colonies into one of the five categories. Colonies were then scored based on GFP expression. The proportion of each colony type within the GFP positive and GFP negative population was calculated and compared to the proportions in the non-transduced control.

For the mock construct, the proportion of colony types was the same between GFP positive and negative colonies, and did not differ from the non-transduced controls. This demonstrates that the effects which are observed within Lin- MNCs transduced with the *CEBPA* constructs are due to the expression of the mutated protein (**Figure 5.7**).

GFP expression correlates with mutant *CEBPA*

To ensure that the colonies which appear as GFP negative under the microscope do not express the mutant *CEBPA* gene, both GFP positive and



	G	M	GM	BFU-E	CFU-E
Mock GFP+	0.0%	2.3%	25.6%	69.8%	2.3%
Mock GFP-	1.8%	4.0%	22.7%	70.1%	1.4%
NT	5.2%	4.9%	30.1%	55.4%	4.5%

Figure 5.7: Comparison of Lin- MNC transduced with mock virus to non-transduced cells

To ensure that the affect observed within the assay was not due to GFP expression or due to the virus construct itself, transduction of the Lin- MNC cells was performed with a virus construct containing only IRES-eGFP. The proportion of colony types were the same between transduced and non-transduced cells, indicating that the effects that are observed within the mutant transduced cells are due to the expression of the mutated protein and not due to viral or GFP expression.

negative colonies were picked from the methylcellulose assays. DNA was isolated and tested for the presence of the N-terminal mutation via the specific real-time assay developed and described in Chapter 4. Real time detection of β2M was used as a control (**Figure 5.8**). All colonies yielded a signal for β2M (**a**), while only GFP positive colonies tested positive for the *CEBPA* mutation (**b**).

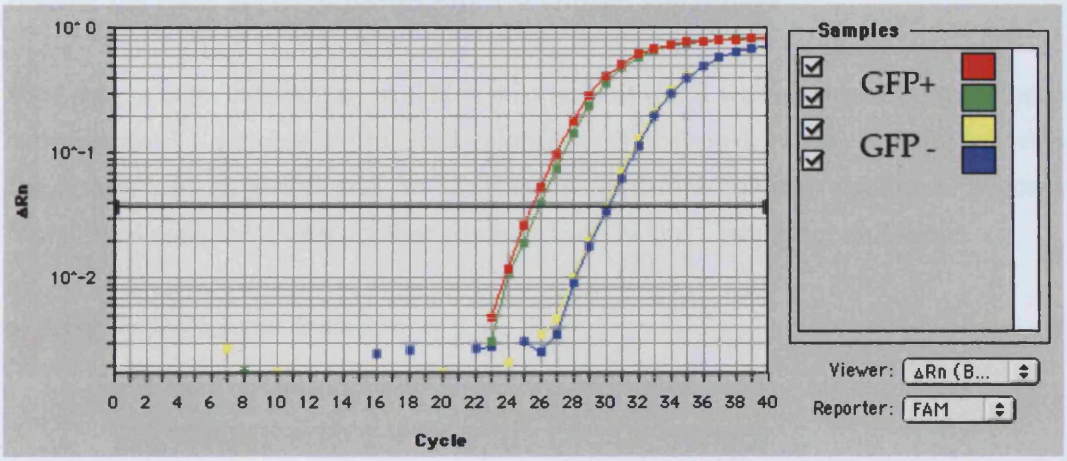
These experiments demonstrate that any effects observed as a result of transduction with the lentiviral constructs are due to expression of the mutant forms of *CEBPA*.

ANALYSIS OF *IN VITRO* ASSAYS

The experiment was performed 4 times on separate occasions, resulting in the assays 1-4 reported in this section of the study. The results are tabulated in **Figure 5.9** and show that both the "N" and "NC" constructs have an effect on the proportion of colony types. Cells successfully transduced (GFP positive) with the "N" construct resulted in a marked increase in the proportion of myeloid colonies and the "NC" construct resulted in an increase in proportion of erythroid colonies.

Analysis of the mutant constructs was performed by calculating the proportion of the five colony types counted present for both GFP positive and GFP negative colonies for each construct. Proportions of colony types were utilised rather than absolute numbers to facilitate comparison between

a) β2M



b) CEBPA

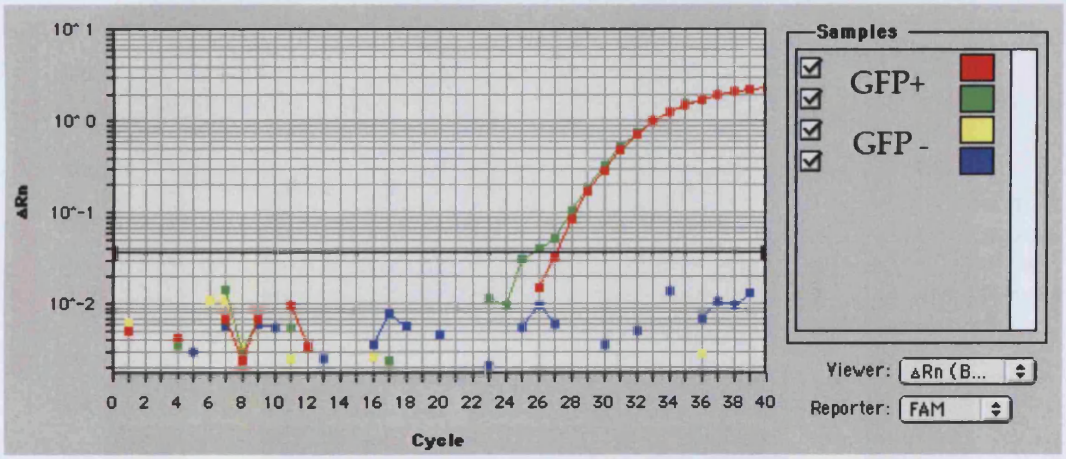


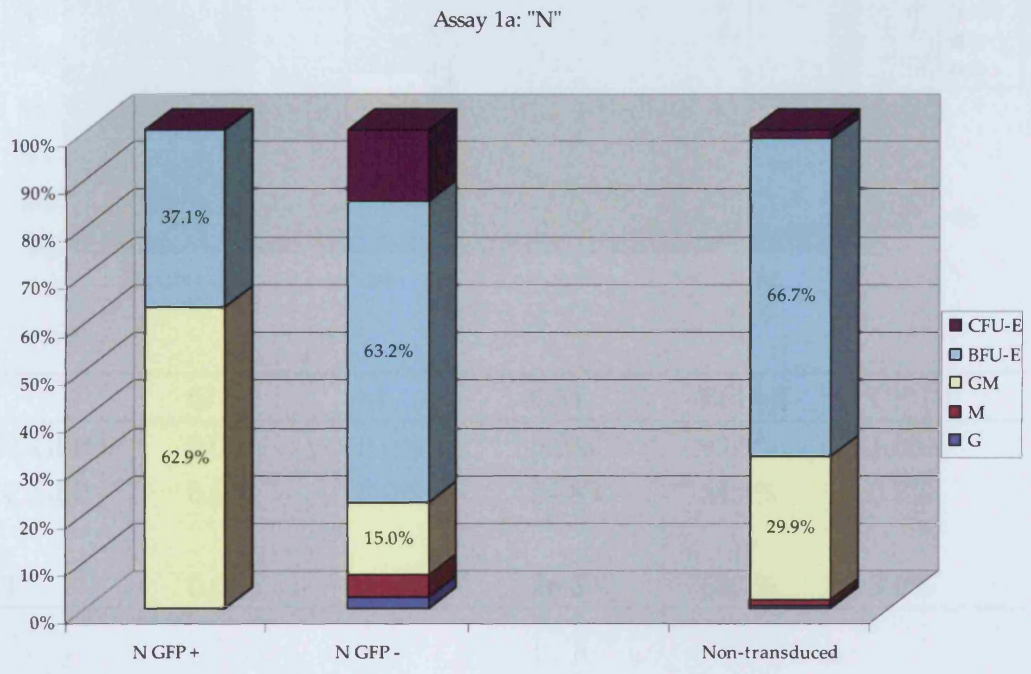
Figure 5.8: Real-time PCR results show that colonies visualised as GFP positive also express mutant forms of *CEBPA*

To ensure that colonies that appeared to be GFP negative via microscopy did not express the mutant *CEBPA* gene, both GFP+ and GFP- colonies were picked from the methylcellulose assays. DNA was isolated and tested for the presence of the N-terminal mutation via the specific real-time assay developed and described in Chapter 4 and with real time against β2M as control. While all colonies resulted in a signal for β2M (a), only GFP positive colonies contained the *CEBPA* mutation (b).

Figure 5.9: Analysis of *in vitro* experiments with mutant *CEBPA* constructs; results for each set of experiments are shown separately

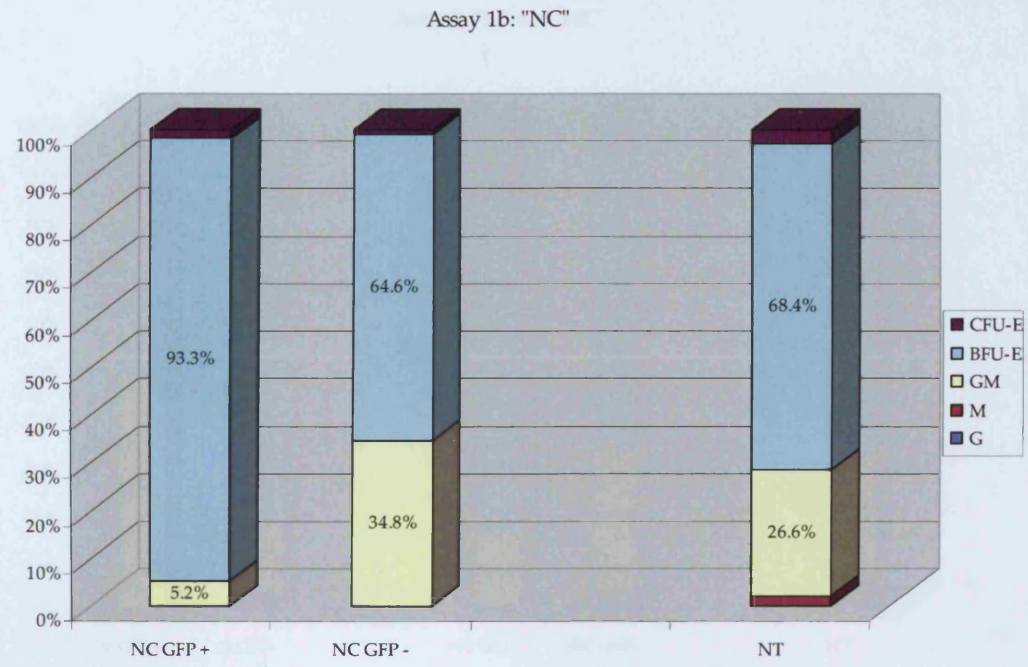
Four assays were carried out in which a consistent trend was observed between each experiment. An increase in myeloid colonies was observed in cells transduced with the N-terminal mutation and an increase in erythroid colonies produced by cells transduced with the construct that contains both N and C-terminal mutations.

a) Assay 1: N-terminal mutation



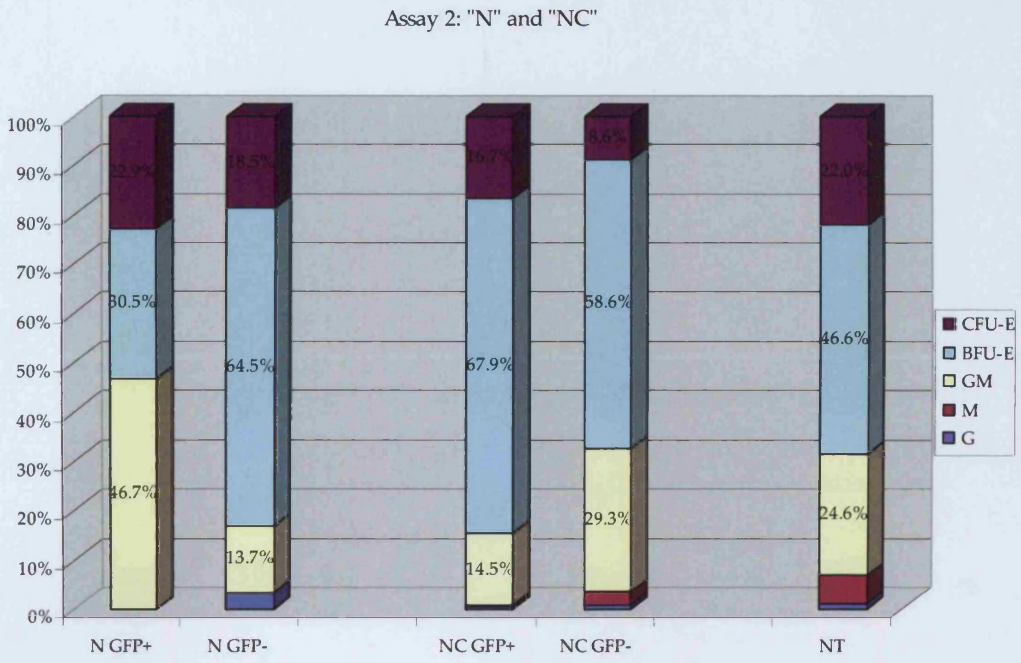
	G	M	GM	BFU-E	CFU-E
N GFP +	0.0%	0.0%	62.9%	37.1%	0.0%
N GFP -	2.1%	4.8%	15.0%	63.2%	14.8%
Non-transduced	0.4%	1.3%	29.9%	66.7%	1.6%

b) Assay 1: N and C-terminal mutation



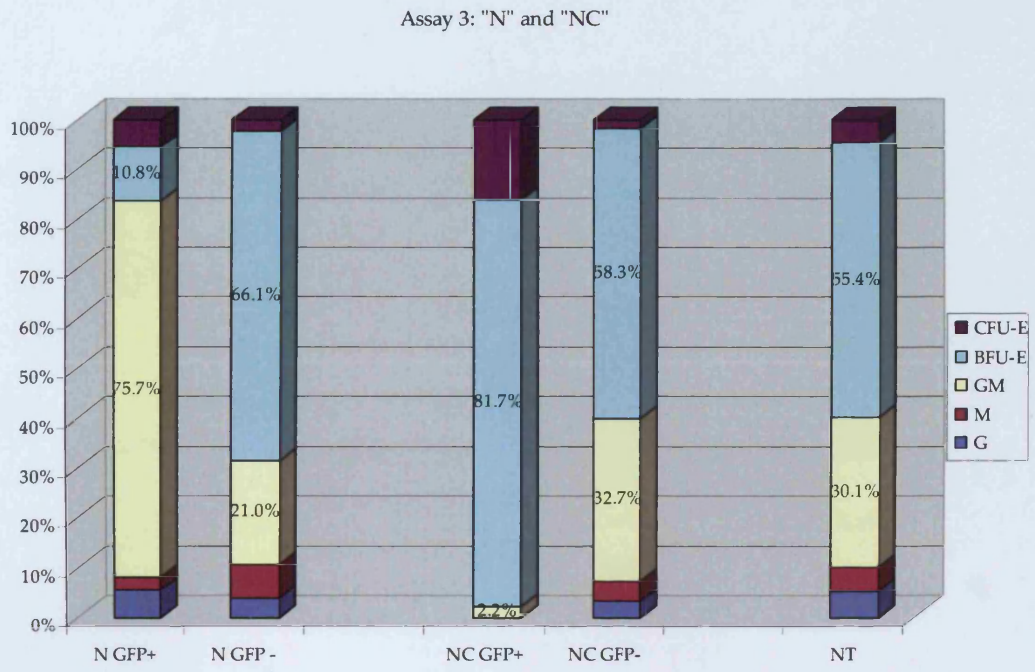
	G	M	GM	BFU-E	CFU-E
NC GFP +	0.0%	0.0%	5.2%	93.3%	1.5%
NC GFP -	0.0%	0.0%	34.8%	64.6%	0.7%
NT	0.0%	2.1%	26.6%	68.4%	3.0%

c) Assay 2: N and NC mutation



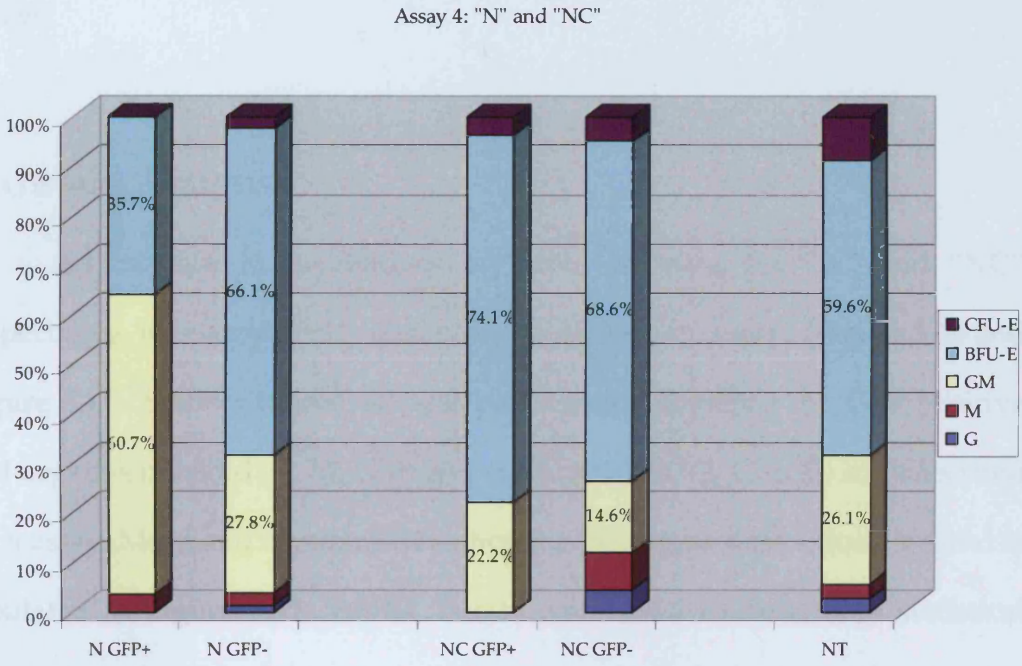
	G	M	GM	BFU-E	CFU-E
N GFP+	0.0%	0.0%	46.7%	30.5%	22.9%
N GFP-	3.2%	0.0%	13.7%	64.5%	18.5%
NC GFP+	0.4%	0.4%	14.5%	67.9%	16.7%
NC GFP-	0.9%	2.6%	29.3%	58.6%	8.6%
NT	1.1%	5.7%	24.6%	46.6%	22.0%

d) Assay 3: N and NC mutation



	G	M	GM	BFU-E	CFU-E
N GFP+	5.4%	2.7%	75.7%	10.8%	5.4%
N GFP -	4.0%	6.7%	21.0%	66.1%	2.3%
NC GFP+	0.0%	0.0%	2.2%	81.7%	16.1%
NC GFP-	3.2%	4.1%	32.7%	58.3%	1.7%
NT	5.2%	4.9%	30.1%	55.4%	4.5%

e) Assay 4: N and NC mutation



	G	M	GM	BFU-E	CFU-E
N GFP+	0.0%	3.6%	60.7%	35.7%	0.0%
N GFP-	1.4%	2.5%	27.8%	66.1%	2.2%
NC GFP+	0.0%	0.0%	22.2%	74.1%	3.7%
NC GFP-	4.3%	7.6%	14.6%	68.6%	4.9%
NT	2.6%	3.0%	26.1%	59.6%	8.6%

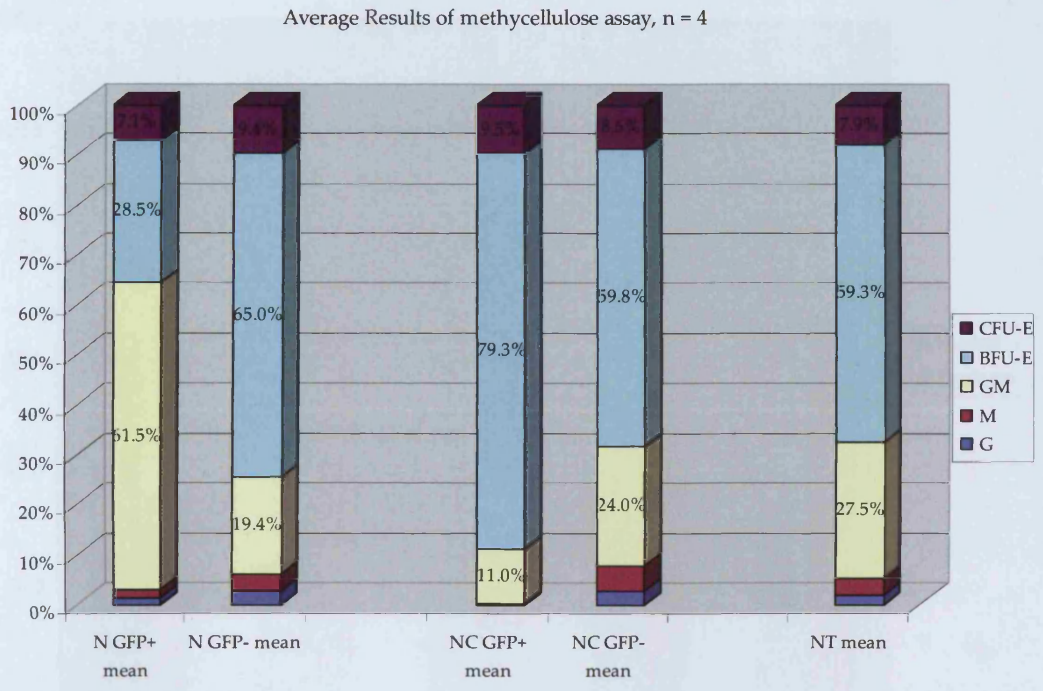
experiments and to account for variation in transduction efficiency between assays.

Statistical Analysis

An increase in myeloid or erythroid colonies, for "N" and "NC" respectively, was consistently observed across the four assays (**Figure 5.10** and **Figure 5.11**). Statistical analysis was performed comparing the GFP positive and negative myeloid (G, M, GM) and erythroid (BFU-E, CFU-E) colonies from each assay. Mean and standard deviation for the assays were calculated and is tabulated in **Figure 5.11**. Paired T-tests were used to determine statistical significance between the proportion of myeloid and erythroid colonies between transduced and non transduced cells (**Table 5.1**).

For "N", the difference in proportion of myeloid and erythroid colonies is statistically significant between the colonies transduced with the mutant construct and both the non-transduced control cells (myeloid $p=0.013$, erythroid $p=0.014$) and GFP negative cells within the assay (myeloid $p=0.0044$, erythroid $p=0.0046$). There was no statistical difference between GFP negative colonies and non-transduced controls (myeloid and erythroid $p>0.05$).

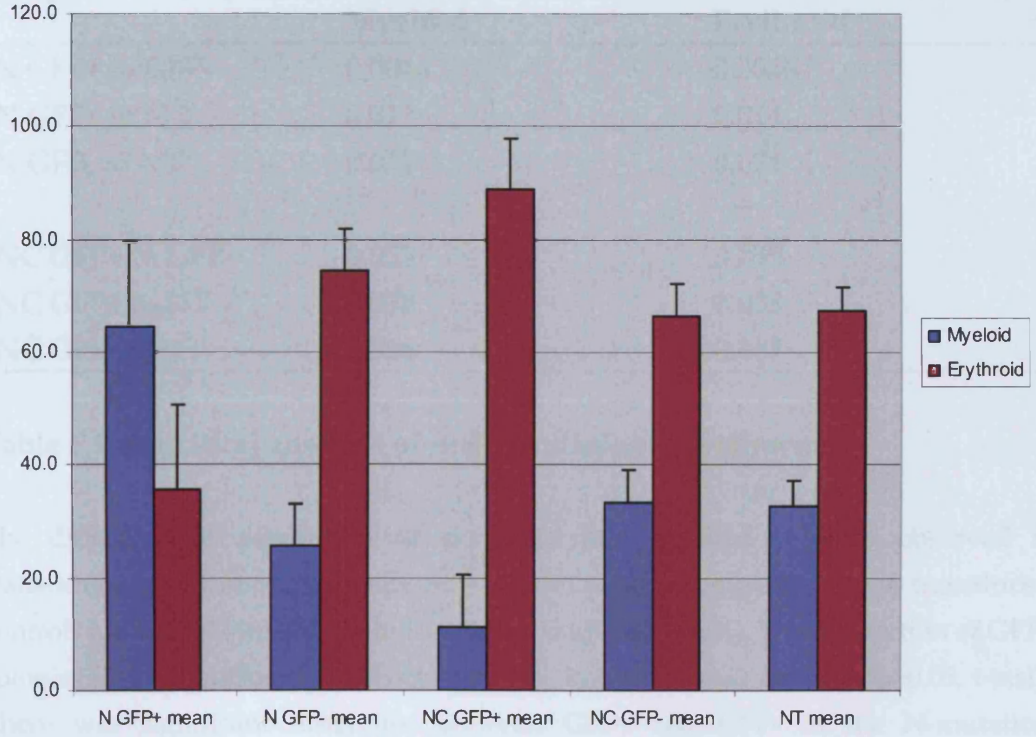
For "NC", the difference in proportion of colony types is significant between the GFP positive transduced colonies and the non-transduced control (myeloid and erythroid $p=0.038$). The GFP negative colonies were not statistically significantly different in proportion to the non-transduced control



	G	M	GM	BFU-E	CFU-E
N GFP+ mean	1.4%	1.6%	61.5%	28.5%	7.1%
N GFP- mean	2.7%	3.5%	19.4%	65.0%	9.4%
NC GFP+ mean	0.1%	0.1%	11.0%	79.3%	9.5%
NC GFP- mean	2.7%	5.1%	24.0%	59.8%	8.6%
NT mean	1.9%	3.4%	27.5%	59.3%	7.9%

Figure 5.10: Combination of methylcellulose results obtained from assays 1-4

The observation of a consistent trend across the assays indicates that the results are unlikely to be due to random insertion events of the lentivirus.



	Myeloid (% total)	Erythroid (% total)
N GFP+ mean	64.4	35.6
N GFP+ sd	15.2	15.2
N GFP- mean	25.6	74.4
N GFP- sd	7.4	7.3
NC GFP+ mean	11.2	88.8
NC GFP+ sd	9.2	9.2
NC GFP- mean	33.5	66.5
NC GFP- sd	5.6	5.6
NT mean	32.7	67.3
NT sd	4.4	4.3

Figure 5.11: Mean of transduced myeloid and erythroid colonies

To simplify analysis colonies were pooled into either myeloid lineage (G, M and GM) or erythroid (BFU-E, CFU-E). The mean and standard deviation are calculated across the 4 assays performed.

	Myeloid	Erythroid
N GFP+ to GFP-	0.0044	0.0046
N GFP+ to NT	0.013	0.014
N GFP- to NT	0.073	0.071
NC GFP+ to GFP-	0.055	0.055
NC GFP+ to NT	0.038	0.038
NC GFP- to NT	0.836	0.837

Table 5.1: Statistical analysis of methylcellulose experiments

The difference in proportion of erythroid and myeloid colonies observed is statistically significant for transduced colonies (GFP+) compared to non-transduced controls for both N and NC mutant constructs ($p<0.05$, t-test). The proportion of GFP- colonies is not significantly different from the non-tranduced controls ($p>0.05$, t-test). There was significant difference between GFP+ and GFP- in the N-mutation construct and a strong trend for the NC construct.

(myeloid and erythroid $p>0.05$). Although it did not reach significance, there was a strong trend ($p=0.055$) for an increase in erythroid colonies upon transduction with the "NC" between GFP positive and negative cells.

MUTANT C/EBPA DOES NOT APPEAR TO AFFECT PROLIFERATION

The total number of colonies per plate for each assay was compared to determine whether expression of the mutant proteins significantly altered the ability of the cells to form colonies within the methylcellulose assays (**Figure 5.12**). Although absolute numbers of colonies per plate varied between assays, there was no difference in colony formation for either "N" or "NC" transduced plates compared to the Nt plates.

The percentage of GFP positive cells was measured 1-3 days following transduction with the lentivirus constructs. This was compared to final percentage of GFP colonies observed to determine whether an outgrowth of transduced cells occurred during the experiments. In general there was a trend for an increase in percent of GFP positive colonies for "NC" constructs and a decrease for "N"; however this was not consistently observed across all experiments (**Figure 5.13**).

TRANSDUCTION WITH WILD TYPE CEBPA

To ensure the changes observed with the transduction of Lin- MNCs with "N" and "NC" constructs was caused by the mutations in gene lentivirus

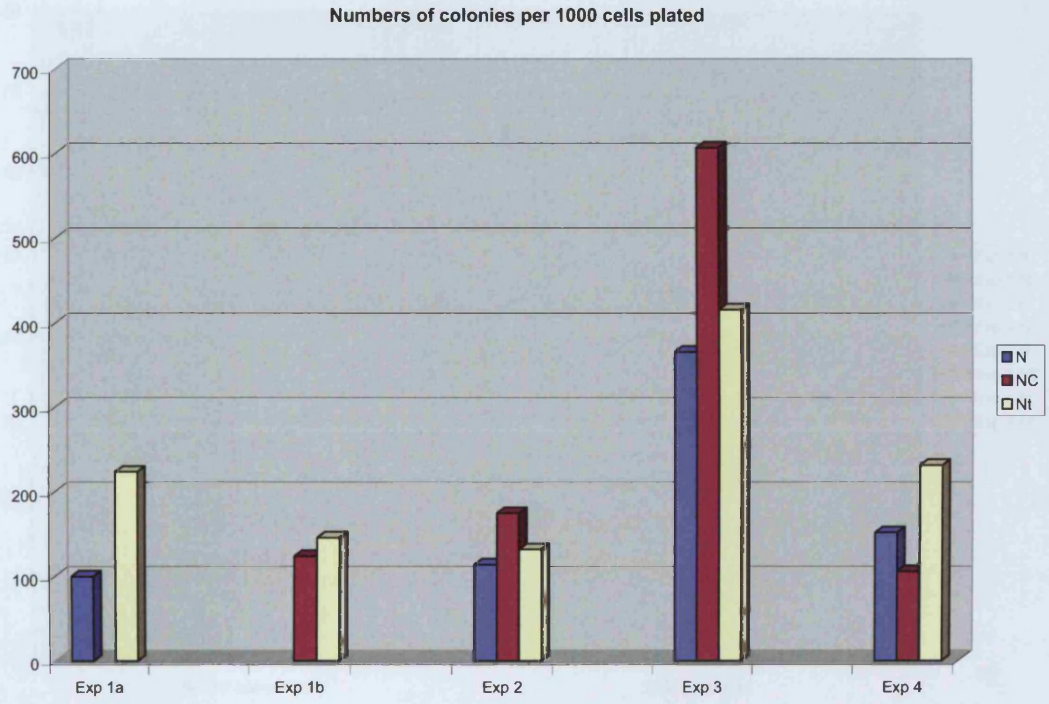


Figure 5.12: Comparison of absolute colony numbers

There is no consistent trend between the numbers of colonies scored on each plate and a particular construct, indicating that variation is not influenced by the particular mutant constructs.

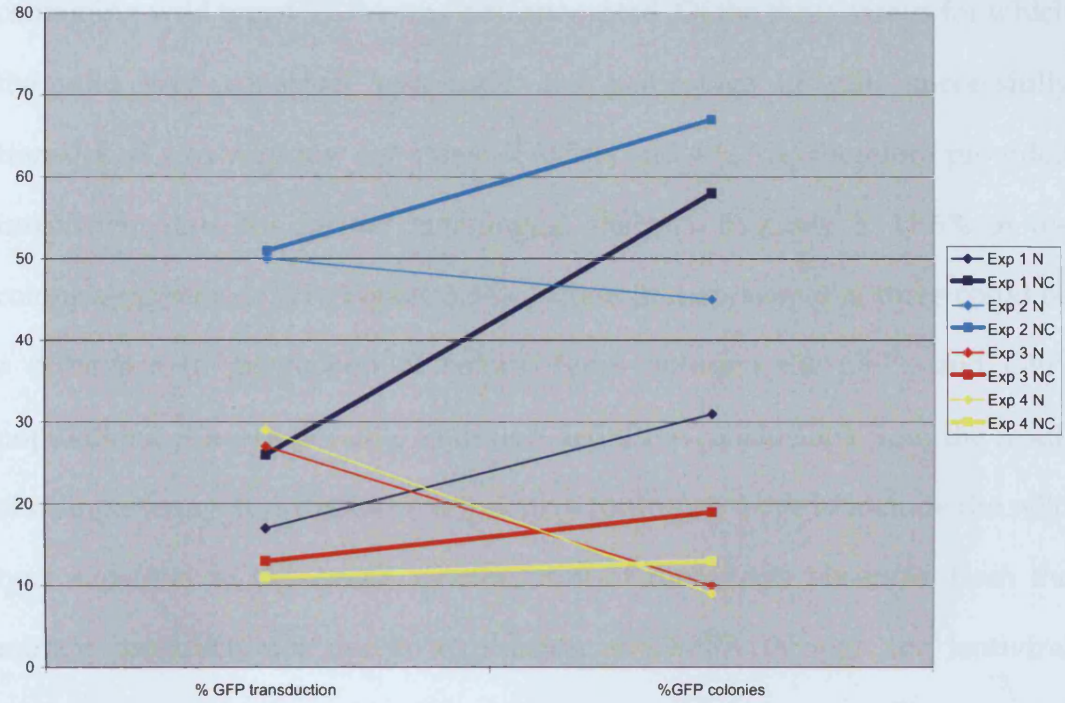


Figure 5.13: Overall there is no significant outgrowth of GFP positive cells for either construct

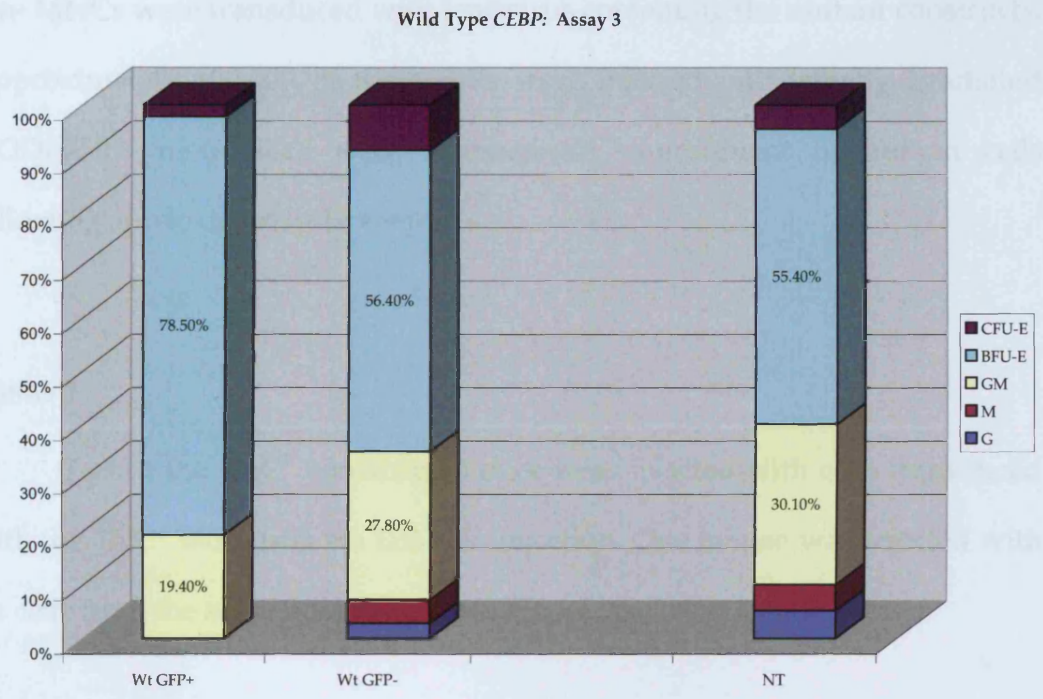
To determine whether the mutant constructs provided a proliferative advantage to the cells they transduced, % GFP was calculated at initial transduction (left hand side of chart) and at the end of the assay (right hand side of chart). "N" constructs are depicted with thin lines, "NC" with thick. Overall, there was no significant outgrowth or depletion of transduced cells for either construct.

containing wild type *CEBPA* was also attempted. Of the three assays for which the wild type construct was used, the percentage of cells successfully transduced was very low for assays 2 (0.5%) and 4 (2.5%) therefore provided insufficient data for further meaningful analysis. In assay 3, 11.6% of the colonies expressed GFP (**Figure 5.14**). Although it appears that there could be a difference in proportion of colony types between the GFP+ and GFP- populations, it is not possible to drawn definitive conclusions from the result of a single assay. It is therefore imperative for future work to include the wild type construct to determine whether any of the effects observed from the mutant constructs are due to expression of *CEBPA* through the lentiviral system rather than from the effect of the mutations themselves. These results however suggest that the effects observed are not only due to general *CEBPA* expression since the "N" and "NC" constructs resulted in different outcomes.

IN VIVO: LONG TERM REPOPULATING ABILITY

The short term *in vitro* methylcellulose assays indicate that differentiation of Lin- MNC is influenced by expression of mutant forms of *CEBPA*. To examine whether the mutant genes have any affect on long-term repopulating cells, NOD/SCID assays were used to examine self renewal ability of transduced Lin- MNCs.

Two *in vivo* experiments were carried out to test the mutant forms of *CEBPA*. These assays correspond to Assay 1 and 2 in the *in vitro* study, with the same transduced cells being used for both assays. As previously described,



	G	M	GM	BFU-E	CFU-E
Wt GFP+	0.00%	0.00%	19.40%	78.50%	2.20%
Wt GFP-	2.80%	4.40%	27.80%	56.40%	8.50%
NT	5.20%	4.90%	30.10%	55.40%	4.50%

Figure 5.14: Transduction of Lin- MNC with wild type *CEBPA*

The lentiviral expression of wild type *CEBPA* could account for some of the effects observed from the assays involving the mutant constructs. Unfortunately transduction efficiencies with the wild type construct were quite poor with only Assay 3 having sufficient GFP+ colonies for analysis. Although there appears that there could be a difference in proportion of colony types between the GFP+ and GFP- populations, it is not possible to drawn definitive conclusions from the result of a single assay.

Lin- MNCs were transduced with lentivirus containing the mutant constructs. Approximately 100,000 of these cells were injected into lethally irradiated NOD/SCID mice. Mice were assessed for engraftment of human cells following approximately 12 weeks.

ASSAY 1

To test the "NC" construct, 5 mice were injected with cells transduced with the "NC" lentivirus via tail vein injection. One mouse was injected with Nt cells from the same pool of Lin- MNCs.

To test the "N" construct, 3 mice were injected with Lin- MNCs transduced with the "N" construct. One mouse was injected with Nt cells from the same pool of Lin- MNCs.

Transduction efficiency of Lin- CB MNCs

Following injection of transduced cells, the remaining cells were returned to media with additional cytokines. These cells were maintained for 1-3 days and then analysed by flow cytometry for GFP expression to determine the efficiency of transduction. In the first assay, "NC" virus resulted in 21% GFP positive cells and "N" virus yielded 17% of GFP positive cells.

Engraftment of "NC" construct (Assay 1a)

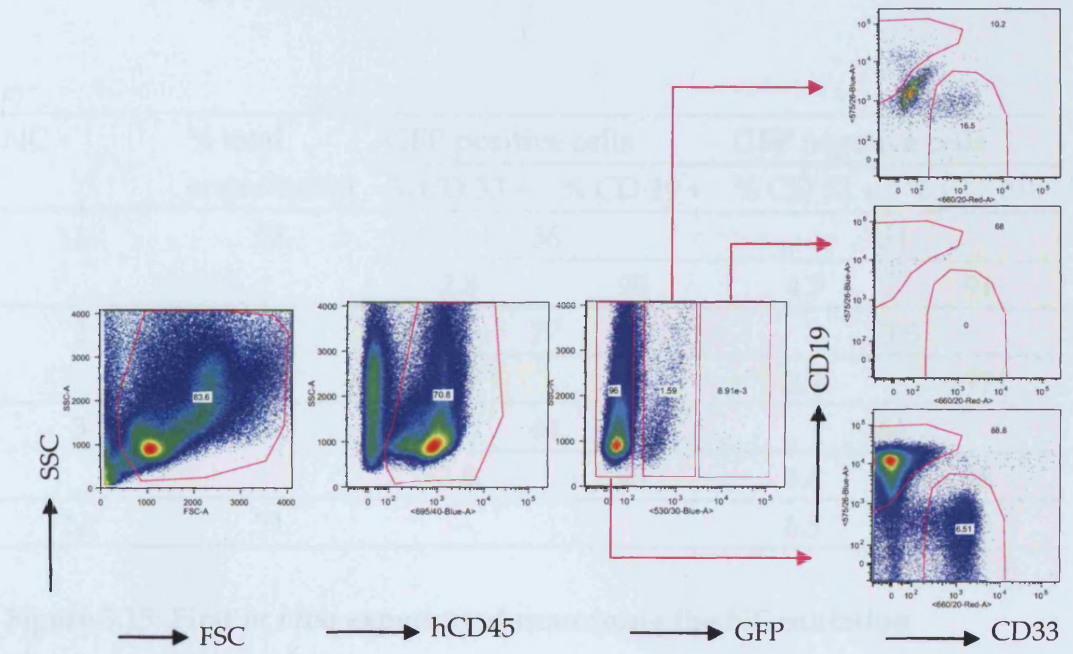
Two of the mice that were injected with transduced cells died before the

end of experiment due to illness unrelated to the experiment. The remaining 3 "NC" mice and the 1 Nt control were analysed for engraftment of human cells by flow cytometry. Antibody against human CD45 was used to test for overall engraftment of human cells within the mouse system. Myeloid and lymphoid engraftment was tested by using CD33 and CD19 markers respectively. Overall human engraftment ranged from 51 to 72% in the transduced cells and was 71% for the non-transduced control (**Figure 5.15**).

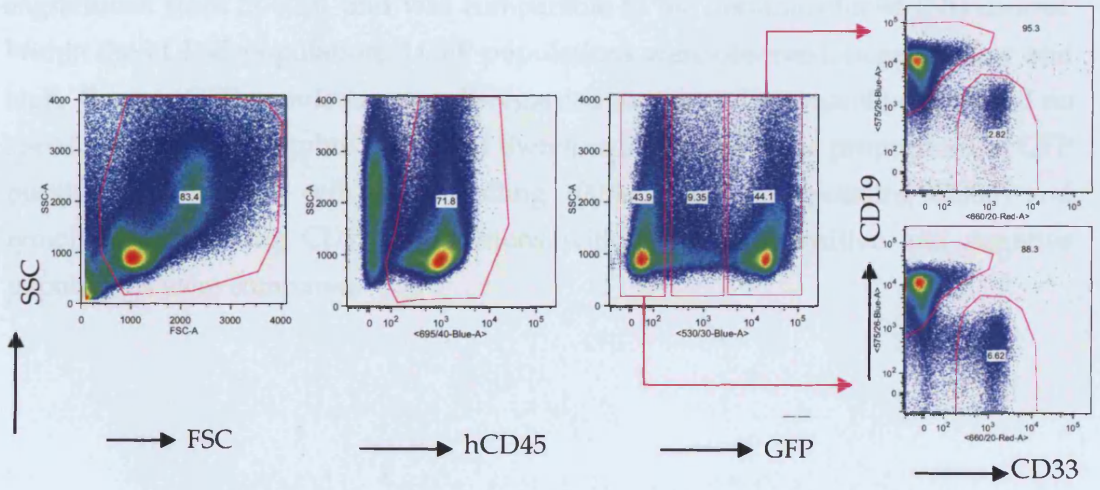
Three distinctive populations of GFP expression were observed, no GFP expression (negative), low GFP expression and high GFP expression. The cells in the low GFP expression were also observed within the non-transduced controls and when examined on a CD19 vs CD33 plot the GFP low population, appear to consist of debris and not fall into populations of a specific cell type. Therefore, this population of cells was not classified as being GFP positive, only those with high expression of GFP (**Figure 5.15**). The amount of GFP positive cells in the transduced populations varied from 36 to 77% of all human cells expressing GFP.

Within the human GFP positive and negative populations, the cells were analysed for myeloid and lymphoid lineage. On average 3% of GFP positive cells expressed CD33 (myeloid marker) and 95% expressed CD19 (lymphoid marker). GFP negative cells from these mice showed on average 4.5% CD33+ and 90% CD19+. The Nt control had 6.5% CD33+ cells and 89% CD19+. A slight reduction in the percentage of GFP positive myeloid cells was observed compared to GFP negative cells; however there was not enough

a) Non-transduced control



b) N and C-terminal mutation



c)

NC	% total engraftment	GFP positive cells		GFP negative cells	
		% CD 33 +	% CD 19 +	% CD 33 +	% CD 19 +
1	51	36		54	
		2.8	95	4.5	91
2	64	77		13.5	
		3	95	2.4	92
3	72	44		44	
		2.8	95	6.6	88
Nt	71	-	-	6.5	89

Figure 5.15: First *in vivo* experiment examining the NC mutation

Engraftment of transduced Lin- MNCs was analysed as described in chapter 2. Overall engraftment was determined by percentage of cells expressing human CD45 within the murine bone marrow. Cells transduced with the "NC" construct ranged in engraftment from 51-72%, and was comparable to the non-transduced (Nt) control. Within the hCD45 population, 3 GFP populations were observed, negative, low and high. The low GFP population was disregarded as it contained mainly debris and no specific myeloid or lymphoid cells and therefore accounts for the proportions of GFP positive and negative cells not equalling 100%. Myeloid (expressing CD33) and lymphoid (expressing CD19) proportions within the GFP positive and negative populations were compared.

information to draw definitive conclusions.

Engraftment of "N" construct (Assay 1b)

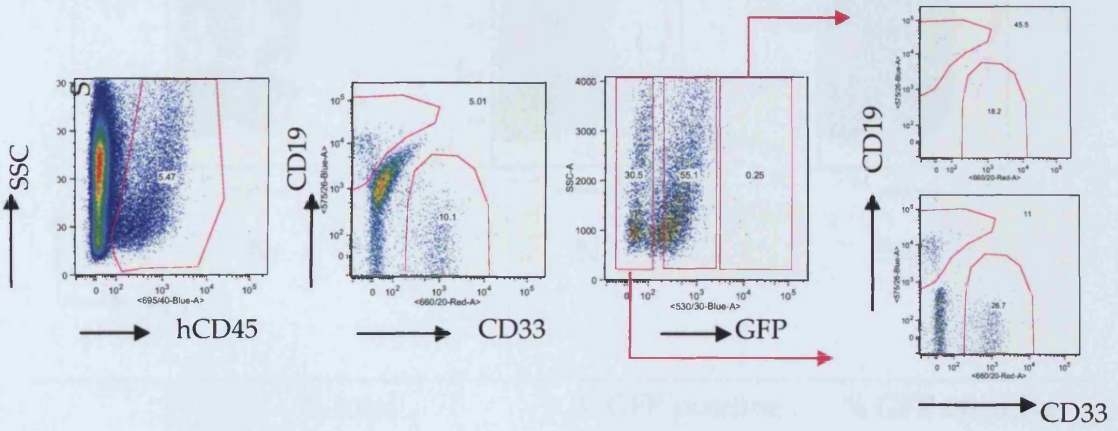
The 3 transduced mice and the 1 Nt control were analysed as described for the "NC" transduced cells. Overall engraftment was lower than the previous "NC" experiment, ranging from 2.6 - 6.1%. Engraftment of the Nt control was 16.3%. GFP positive engraftment was below 1% of engrafted cells and therefore of insufficient number for further analysis (**Figure 5.16**).

ASSAY 2

The *in vivo* experiment was repeated, with Lin- MNCs from the same preparation transduced with both "NC" and "N" constructs along with an Nt control. Four mice were injected with cells transduced with "N", 2 mice for "NC" and 1 mouse for the Nt control. Mice were sacrificed at 11 weeks and analysed for engraftment as described previously for assay 1 (**Figure 5.17**).

As in the previous experiment, overall engraftment was greatly reduced in the "N" cells (1.4 - 5.1%) as compared to the Nt control (29.8%). Engraftment of GFP positive cells was observed only in one of the four mice (N3, 0.34%). Engraftment of the "NC" cells was varied. The first mouse had 13.9% human engraftment, with 4% GFP positive cells. The second mouse had 59.8% engraftment with 28% of the cells being GFP positive. Of the GFP positive cells 18.4% were myeloid and 79.8% were lymphoid and showed a

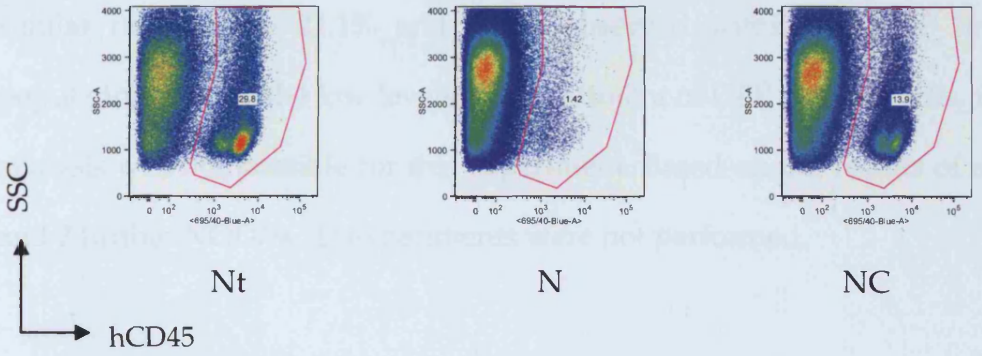
a) N-terminal mutation



N	% total engraftment	GFP positive cells		GFP negative cells	
		% CD 33+	% CD 19 +	% CD 33 +	% CD 19 +
1	5.5	0.25		30.5	
		-	-	11	27
2	2.6	0.1		10.6	
		-	-	1.6	14
3	6.1	0.04		67.1	
		-	-	9	69
Nt	16.3	-	-	6	85

Figure 5.16: First *in vivo* experiment: N-mutation

Engraftment of transduced Lin- MNCs was analysed as described in chapter 2. Overall engraftment was determined by percentage of cells expressing human CD45 within the murine bone marrow. Cells transduced with the "N" construct ranged in engraftment from 2.6-6.1%, lower than the non-transduced (Nt) control. There was very little GFP positive engraftment within the "N" construct, preventing further analysis.



	% total engraftment	% GFP positive cells	% GFP negative cells
NC 1	13.9	4	11.4
NC 2	59.8	28	57.5
N 1	1.4	-	19.4
N 2	5.1	-	62.7
N 3	1.4	0.34	12.3
N 4	2.1	-	3.55
Nt	29.8	-	82.5

Figure 5.17: Engraftment results corresponding to the second *in vivo* experiment.

Engraftment of transduced Lin- MNCs was analysed as described in chapter 2. Overall engraftment was determined by percentage of cells expressing human CD45 within the murine bone marrow. Cells transduced with the "NC" construct ranged in engraftment from 13.9-59.8%, comparable to the non-transduced (Nt) control. Cells transduced with the "N" construct had lower engraftment as compared to the control. Similar to the previous experiment there was little GFP positive engraftment, preventing further analysis.

similar ratio to the 21.1% and 76.8% observed within the GFP negative population. Due to the low levels of engraftment of GFP positive cells, further analysis was not possible for this experiment. Based on the results of assay 1 and 2 further NOD/SCID experiments were not performed.

SUMMARY

This study has started to test the contribution of mutant forms of *CEBPA* on haematopoietic development. To accomplish this, Lin- MNCs were transduced with lentiviral constructs containing mutant forms of *CEBPA* and assessed for changes in short term repopulation ability and differentiation by *in vitro* methylcellulose assays and for long term repopulating ability by *in vivo* xenotransplant models. The results of the *in vitro* experiments indicate that the mutant forms of *CEBPA* play an important role in determining the lineage fate of Lin- MNCs.

POSSIBLE MOLECULAR MECHANISM FOR CHANGE IN COLONY TYPES DUE TO MUTANT FORMS OF C/EBPA

CEBPA has been shown to play an important role in myeloid lineage development and it is therefore not surprising that mutant forms of the protein are observed to have an effect on differentiation of Lin- MNCs. Methylcellulose assays primarily examine early progenitors, cells with limited self-renewal capacity and it is important to note that the colonies scored do not consist of terminally differentiated cells.

Both the "N" and "NC" constructs appear to function in a dominant negative manner against wild type C/EBPα since no attempt was made to silence the endogenously expressed protein. To perform its function as a transcription factor, C/EBPα must dimerise before it is able to bind to its DNA recognition sequence. It can either form homodimers or heterodimers with other C/EBP family members. In the case of the 30kDa isoform (which corresponds to the "N" mutation construct), the protein retains its dimerisation and DNA binding capabilities, however it has reduced transcriptional activation ability due to the lack of the first transactivating domain[65].

In the *in vitro* experiments presented in this chapter, expression of the "N" construct in Lin- MNCs was shown to promote colony differentiation down the myeloid lineage. It is possible that this isoform retains sufficient functional capacity to initiate myeloid differentiation, but below a threshold level to allow terminal differentiation of the cells, which was not assessable by this assay, resulting in an accumulation of myeloid progenitors. Furthermore it is possible for other members of the C/EBP family help to compensate for the lack of functional C/EBPα and allow haematopoiesis to proceed.

As discussed in the previous chapter, mutations within the C-terminal region of C/EBPα tend to cluster together. The majority of internal tandem duplications and in-frame duplications fall within the small region separating the DNA-binding basic domain from the leucine zipper dimerisation domain.

Figure 5.18 shows the location of mutations for the 4 patients studied in this thesis along with the DNA binding and dimerisation domains of the protein.

It is possible that in the cells expressing the mutant proteins that the dimerisation function is maintained and therefore endogenous wild type C/EBPα is sequestered. However, the altered protein structure in the adjoining region results in the dimerised protein losing DNA binding ability. Due to the alpha helical nature of the protein in this region, even small insertions are sufficent to disrupt its binding ability and therefore cause the loss of function as a transcription factor and its ability to promote myeloid development[198]. As is shown previously in **Figure 1.9**, C/EBPα activity is upregulated during the differentiation process from the CMP to the more committed GMP population. Therefore, a loss of function for C/EBPα may result in cells differentiating towards the erythroid lineage due to a block in myeloid differentiation, rather than promoting erythroid differentiation.

The "NC" mutant construct did not completely abolish myeloid colony formation within the system. Although the mutant isoforms may function in a dominant negative manner by sequestering the wild type protein, is it likely that sufficent endogenous C/EBPα remained in a functional state to produce the myeloid colonies observed.

There is ample evidence that expression of *CEBPA* plays a crucial role in myeloid development. Murine models have demonstrated that enforced expression of *CEBPA* in HSCs results in the loss of erythroid progenitors and

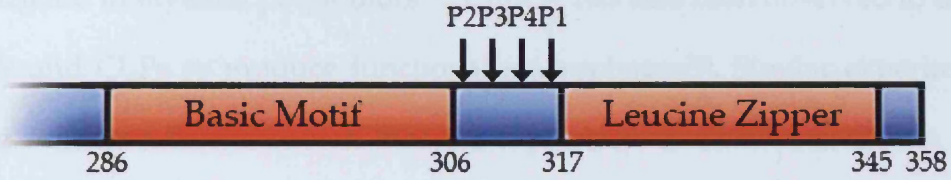


Figure 5.18: Mutations in *CEBPA* cluster in the region between the basic motif and leucine zipper

The C-terminal domain of C/EBPα is involved in DNA binding and dimerisation through its basic region and leucine zipper, respectively. In the four patients studied for mutations in CEBPA, all the C-terminal insertion mutations fall in between the basic region and the leucine zipper. The insertion of amino acids disrupts the alpha helical secondary structure of this region. In the case of the mutations between the regions, it is possible that the protein is able to retain it dimerisation ability, but loses its ability to bind to its target DNA sequence.

an increase in myeloid progenitors[199]. CEBPA has also been observed to direct MEPs and CLPs to produce functional macrophages[200]. Similar experiments using human cells has shown that the expression of CEBPA in CD34+ cells leads to granulocytic differentiation and inhibits erthyroid differentiation[201].

Schwieger *et al*[135] demonstrated that expression of the 30kDa isoform in human CD34+ cells isolated from cord blood inhibits the differentiation of both myeloid and erthyroid lineages. While this is not consistent with our results (overall number of colonies not reduced) this can be explained by the difference in methodology where we analysed of immature cell types rather than examining fully differentiated cells. In keeping with our results using the "NC" contruct, Liu *et al*[136] demonstrated that primitive erythropoiesis is stimulated in zebrafish upon expression of the 30kDa isoform.

Furthermore, in familial cases of AML involving *CEBPA* mutations, patients are able to live for many years with a germline "N" mutation, indicating that the effects of a single mutation is necessary but not sufficient for leukaemogenesis, which is in accordance with our observed effects. The "N" construct allows for myelopoiesis, where the addition of the C-terminal mutation results in a block in this differentiation pathway.

ENGRAFTMENT ASSAYS

A major limitation of the *in vivo* experiments was the quality of lentiviral production. Engraftment of GFP positive cells was low for cells

transduced with the "N" construct and overall engraftment was lower for these cells than for the non-transduced controls and cells transduced with the "NC" construct (**Figure 5.19**). It is likely that the reduction in levels of engraftment from the cells transduced with "N" construct is due to the virus preparation itself since GFP negative (non-transduced) cells within the transduction assay also exhibited reduced engraftment potential. While the same viral preparation was used in our parallel *in vitro* experiments, fewer cells and therefore less virus was required for these experiments.

To determine whether viral toxicity played a factor in the reduced engraftment of the "N" construct another preparation of the lentivirus should be prepared and tested. Although it is unlikely that toxicity of GFP would play a role, the transduction of a mock virus construct would help to eliminate that possibility.

For the cells transduced with the "NC" construct, overall engraftment levels were comparable to those of the non-transduced control, although as in the *in vitro* assays transduction efficiency was variable. In Assay 1, no significant change was observed between the lymphoid and myeloid populations of the engrafted cells when comparing transduced to non-transduced cells. However, as the *in vitro* experiments demonstrated, the significant difference is between the proportion of myeloid and erythroid progenitor cells. A limitation of the NOD/SCID model is that to achieve human erythroid engraftment, the mice require supplementation with the

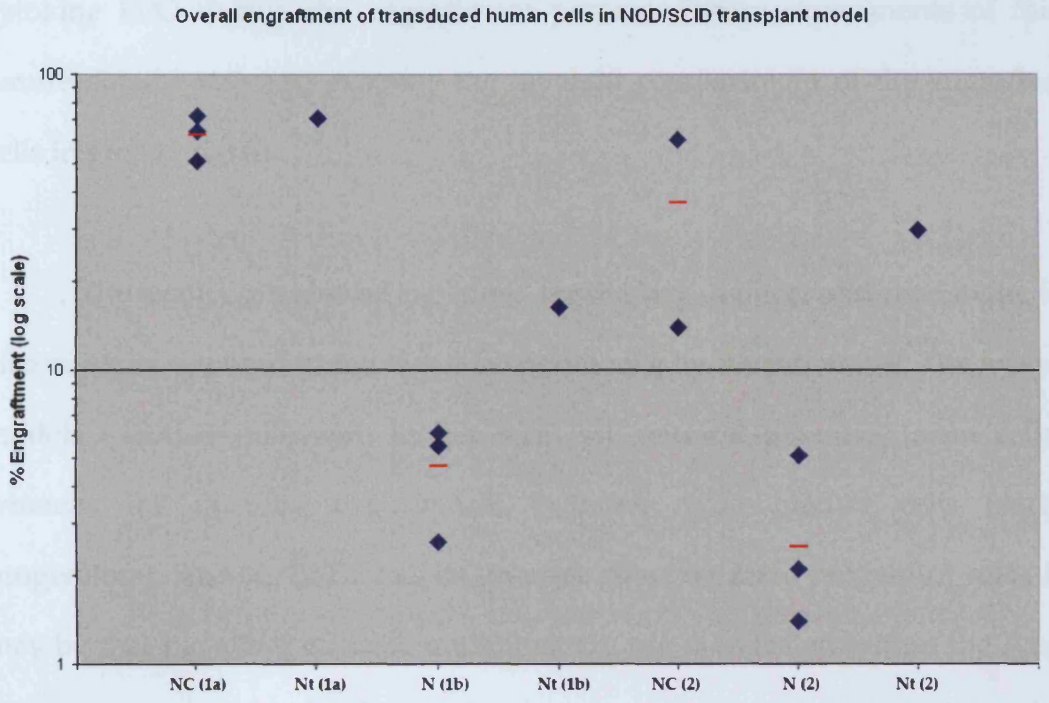


Figure 5.19: Overall engraftment for *in vivo* experiments

Each mouse is represented by a blue diamond. Mean % engraftment is represented by a red bar. A reduction in overall engraftment (both GFP positive and negative cells) for "N" construct transduced cells indicates potential toxicity in the viral preparation. Although engraftment levels for the "NC" construct varied, the mean engraftment percentages were similar to the Nt controls.

cytokine EPO during the engraftment process. Future experiments of this nature should strive to examine the myeloid compartment of the engrafted cells in greater detail.

The results can also be explained for the lack of effect observed in the *in vivo* cases, as opposed to the *in vitro* experiments, by the target cell. The *in vivo* models examine cells with longer term self-renewal potential (stem cells) whereas the *in vitro* experiments examine more mature cells (early progenitors). Since C/EBPα has its greatest affect on early progenitor cells, it may be that the effect of these mutations are not manifested within the long term repopulating HSCs and therefore not observable by the xenotransplantation models.

FUTURE WORK

The results from the *in vitro* methylcellulose assays demonstrate that it is likely that mutation combinations in *CEBPA* play important roles in myeloid differentiation. The work that has been reported here today raises some interesting questions and further experimentation should be undertaken, such as:

- the examination of the "C" mutation alone
- the examination of both "N" and "C" mutations together (mimicking biallelic mutation distribution)
- inclusion of wild type and mock controls
- silencing of endogenous C/EBPα

- targeting cell types (HSCs and early progenitors)
- further proliferation assays

Since transduction of the same cells with the "N" and "NC" constructs yielded quite different results, the next logical step would be to examine the effect of just the C-terminal mutation. If the effects of the "NC" construct are due to the abolishment of DNA binding ability of the protein, a "C" construct would likely produce a similar effect. Furthermore, although both the N and C-terminal mutations are observed on the same allele in some cases, the more prevalent situation is biallelic mutations, with an N-mutation on one allele and a C-mutation on the other. To emulate this in our assay, both the "N" and "C" construct would be used to transduce cells together to observe the effects of the two mutations in concert.

Furthermore, silencing endogenously expressed *CEBPA* via siRNA targeted to the 3' UTR would eliminate any remaining wild type allele. In work performed by Nerlov's group, they have demonstrated that the expression of the 30kDa isoform of *Cebpa* in a *Cebpa* null background results in the development of overt leukaemia in mice. Silencing endogenous *CEBPA* in human cells and then transforming them with the various mutant constructs could potentially give an indication of whether a similar system would have transformation properties in human cells.

Furthermore, the possibility of targeting specfic cell types for transformation should allow further refinement of the assay system. In a

similar manner to the studies examining the transformation ability of fusion proteins, cells sorted into different cell types, such as HSCs, CMPs and GMPs could be transduced with the *CEBPA* mutant constructs to determine whether the mutation has the same effects in different cell types.

Although the *in vitro* assay gave a good readout of the effects on differentiation of the mutant constructs, possible proliferation effects were not assessed within this assay. To determine whether the expression of the mutant forms of C/EBPα alter the self-renewal capabilities of the cells, serial replating and proliferation assays can be conducted.

As discussed briefly previously, the inclusion of control wild type *CEBPA* is important to ensure that effects observed are due to the action of the mutations in the gene and not merely a product of expression of *CEBPA* via the lentiviral system. However, since different effects were observed for the two mutant constructs it is unlikely that these results are due to the expression of *CEBPA* alone. Poor transduction efficiency was a limiting factor in the case of the wild type control, therefore future work would benefit from adjusting methodology (such as adding a spinoculation step) to improve transduction efficiency.

CHAPTER SIX: DISCUSSION

AML is a clonal, malignant proliferation of undifferentiated haematopoietic cells that is morphologically, genetically and therapeutically heterogeneous. The leukaemic blasts that make up the bulk of the disease have low proliferative potential and are maintained by a rare population of LSCs responsible for propagating the disease. Following completion of the human genome-mapping project, we find ourselves in the fortunate position to be able to interrogate each gene for both mutation and gene expression changes in AML and how these events arise in the LSC population. The success of these approaches is evident from several recent reviews from the Bloomfield group[5, 70] who have compiled these data into a speculative molecular prognostic discriminator (**Figure 1.9**). Satisfyingly, we have made several contributions to these analyses. These have been discussed at the end of each chapter and are summarised below. An important new consideration however is not only the nature of these mutational events but also the order and timing by which these events arise. We have tried therefore to trace how these events might have arisen by analysing or inferring mutational patterns within the initial HSC, the corresponding LSC and more typically myeloblast cell population.

We sought to answer some simple questions relating to the nature, pattern and order by which these mutations arise in NK AML, rather than confining ourselves to merely testing for the presence or absence of mutation.

What types of mutations are present?

How many genes are involved?

What length of time does it take for the disease to develop?

In which order do the mutations arise?

What is the function of mutations in AML?

What types of mutations are present and how many genes are involved?

Mutation screening of a panel of 88 NK AML patients for 8 genes (*FLT3, CEBPA, RUNX1, NRAS, cKIT, PTPN11, NPM1* and *WT1)* allowed us to build a picture of the pattern of gene mutations involved in leukaemogenesis. The majority (85%) of patients harbour mutation in at least one gene, with a large proportion containing multiple mutations. Of these, 44% of the AMLs contained mutations in both Class I (deregulation of proliferation) and Class II (block in differentiation) that supported the hypothesis, put forward by Gilliland and colleagues that deregulation of both pathways are required for the development of overt leukaemia.

Mutations occur in a non-random manner

In keeping with this hypothesis mutations do not arise independently of each other. This is a well described phenomenon in the literature. In our series for example *FLT3*-ITD and mutations in *NRAS* occurred independently of each other. In work published by our group[143], *cKIT* mutations were observed exclusively in favourable risk patients and absent in both

intermediate and poor karyotype. Similarly, mutations in *CEBPA* and *NPM1* occur almost exclusively in intermediate/NK cases. The description by our group[100] of a familial case of AML involving *CEBPA* mutations yields further insight into the process of leukaemogenesis whereby N-terminal mutations occurred in the germ-line and somatic C-terminal mutations were required for the onset of overt disease. Such long latency periods preceding AML in this family indicate the requirement of additional cooperative mutations and that mutations in *CEBPA* are a key event in leukaemogenesis. In contrast, patients with familial Wilms' tumour inheriting a germline *WT1* mutation are not predisposed to the development of AML. This suggests that although *WT1* mutations may play a role in AML, they are not an initiating event in leukaemogenesis.

Since a proportion of our cases (15%) showed no mutation in any of the genes tested we also performed selective mutational studies in 2 candidate genes, *CCND3* and *FES*, based on their cellular function. In keeping with the trend in mutation studies of AML mutation analysis was restricted to particular motifs rather then analysing the entire coding region of the gene. No mutations were found within the TKD of *FES* and a single mutation was observed within exon 5 of *CCND3*. This mutation results in an in frame deletion of 51bp of exon 5 and follows a very similar theme to many of the other mutational events in AML where in frame (*CEBPA, FLT3*) and out of frame (*CEBPA, RUNX1, WT1*) insertions or deletions occur frequently. Despite performing extensive mutation analysis we failed to detect mutations in several cases within our cohort demonstrating that additional genes or other

mechanisms of gene deregulation must exist.

Mutations and prognosis

Although mutations in NK AML can be used to stratify patients within the NK cytogenetic risk group into favourable and poor prognostic cases[5], our panel was not sufficiently large to recapitulate these findings. This is not surprisingly as cases from our centre were selected based on the quantity of frozen material rather than on age or the treatment regimes ascribed to the individual patient. However, given the literature on prognostic relevance of gene specific mutations in NK AML, we would strongly support the introduction of routine mutation screening of commonly mutated genes, particularly *FLT3, CEBPA* and *NPM1* at diagnosis for evaluation as part of the clinical management of patients with this disease.

In what order do mutations arise?

Although the mutation screening of our panel of NK AML patients was able to shed light on the types and numbers of mutations required for the generation of AML, merely examining the myeloblast compartment did not address the order by which these genetic lesions arise. The work presented in this thesis on aUPD acting as a second hit in the generation of homozygous mutations has started to address these points and identify how normally occurring mitotic recombination can contribute to leukaemogenesis. In these studies homozygous mutations of *FLT3, CEBPA, RUNX1* and *WT1* were

identified in 7 cases. Until this point and the work of Bryan Young and colleagues[183, 202] it had been difficult to provide a plausible explanation for increased mutation dosage observed for mutations such as *FLT3*-ITD[71]. This is particularly relevant in AML since patients having a high *FLT3* mutant to wild type ratio have a poorer prognosis than patients with lower mutation loads. The link with aUPD provides a clear explanation for how this arises and discriminates between other possible causes, for example hemizygosity of chromosome 13 or localised deletion of the *FLT3* locus. Regions of aUPD has also been observed in other haematological disorders and solid tumours giving the impression that mitotic recombination leading to homozygosity of a pre-existing lesion plays a role in cancer in general.

It is important to point out that we were not able to determine the basis of daughter selection in all cases (aUPD 6p, 9p, 11p and 11q) studied. Based on our current experience it seems plausible that within these regions of homozygosity there must exist genetic lesions that confer an advantage to the cell and effect clonal outgrowth with the removal of the wild-type allele. aUPD of 6p appears to be a common phenomena in haematological malignancies, including follicular lymphoma[203] and therefore should be an area of focus for future research. *CCND3*, which is located on 6p, was screened for mutations and while a mutation was identified exon 5 within the NK panel, the gene was not mutated in aUPD cases suggesting that the selective basis is located elsewhere on chromosome 6. We are also unable to rule out the possibility that epigenetic changes or shifts in tumour allelotype could be responsible for daughter selection following mitotic recombination. The further refinement of

the technology to provide more targeted information on the regions of homozygosity should be able to narrow the focus of these positional studies further.

Besides looking for mutations in genes known to be involved in other types of cancers, another approach to discovering new mutations within AML would be to conduct systematic genome wide screening on several cancers to elucidate possible key changes. A recent study by Sjöblom *et al*[204] utilised this method of systematic screening in breast and colorectal cancers, finally identifying 189 genes out of the 13,000 they screened as playing a key role in the generation of these diseases. While an advantage of AML over solid tumours is a general lack of chromosomal instability, the absence of known mutations in a subset of patients with NK AML indicate that there are likely to be genetic events contributing to the development of these diseases that have not yet been identified that may benefit from such a systematic screening approach.

The occurrence of homozygous *WT1* mutation has proven particularly important to our group[168]. We have since evaluated the role of *WT1* mutation in NK AML. These mutations occur in approximately 10% of cases with 30-50% of mutations indicative of increased mutation dosage and the occurrence of aUPD. The group is in the process of determining the prognostic relevance of these mutations; early suggestion is that these mutations are associated with failure of induction chemotherapy and a trend towards inferior overall survival[168]. It is of interest that not all cases of aUPD in 11p

involve mutations at the *WT1* locus. We have excluded mutations within the coding region of *PU.1* transcription factor in these cases and are now considering the possibility for mutation or allelotype alteration in an upstream enhancer element as the causative mutation as the underlying selective mechanism. It has been recently shown that a polymorphism in an upstream enhancer element of *PU.1* affects transcription of the gene[205]. As was discussed in the introduction, the concentration of transcription factors plays an important role in the generation of AML and therefore any alterations to the production of these proteins may be the mechanism of leukaemogenesis in this model.

Minimal Residual Disease

Reflecting an inability of current treatments to specifically target the LSC, greater than 50% of patients with NK AML will relapse having achieved complete remission. Unlike in the favourable risk setting where fusion transcripts are routinely used to monitor MRD and are predictive for impending relapse, the absence of a molecular signature in NK patients has limited MRD analysis in these patients. *WT1* gene expression has long been considered a good marker for MRD detection for about 70% of all cases[18]. The identification of *NPM1* mutation probably offers a better strategy in the future for NK patients since it occurs in approximately 60% of cases of NK AML and the group of Falini and colleagues[206] have reported its application for the routine testing of MRD. In the Bart's cases we chose to analyse *CEBPA*, mutations in this gene are less frequent and although clustered to particular

regions of the gene require the development of patient specific assays. In keeping with *NPM1, CEBPA* mutations appear to be an initiating event in leukaemogenesis being present with greater than 90% concordance between diagnostic and relapse samples[119]. In the absence of an appropriate retrospective material we utilised our *CEBPA* mutation specific assays to examine the occurrence of mutation within the leukaemia equivalents of the HSC compartments. We were able to show that CEBPA mutations occurred within all tested compartments which is in agreement with familial cases involving the mutations.

What is the function of mutations in AML?

The gene *CEBPA* is particularly suitable for studies examining the clonal evolution of AML due to its intronless structure, which allows us to examine the allelic distributions of mutations. While commonly the occurrence of 2 mutations within a gene is presupposed to reflect mutations on alternative alleles, in *CEBPA* we did noticed the occurrence of biallelic mutations, but also the co-localisation of mutations on the same allele, a phenomenon also reported by other groups[119].

We made two mutant constructs of the *CEBPA* gene, one containing the N-terminal mutation ("N") and the other containing both mutations ("NC"). Our work was able to demonstrate that Lin- MNCs were driven to either myeloid or erythroid lineage when transduced with "N" and "NC" respectively as compared to non-transduced cells within *in vitro*

methylcellulose assays.

In *CEBPA* it is clear that this transcription factor plays an important role in leukaemogenesis. Evidence from familial cases and from its presence within the LSC indicates that mutations in this gene are an early event however are not enough for the development of frank leukaemia immediately on its own.

Target cell of mutation events in leukaemogenesis and *CEBPA*

As studies have shown that the location of mutations and genetic events can play an important role in AML development it is important to determine when these key events take place. In our study, leukaemia samples were sorted into the LSC, and the equivalent of CMP and GMP compartments in normal haematopoiesis and then tested for presence of *CEBPA* mutations, which were detected in all compartments for all samples. This however, does not address in particular where the mutation may arise if the mutation itself is able to alter the phenotype of the cell. An issue that may arise is whether the leukaemogenic event (ie the mutation in *CEBPA*) dictates the cell phenotype or whether the cell type in which the mutation occurs may influence the phenotype of the abnormality.

As discussed in the introduction, transduction of various cell types with fusion proteins may dictate the phenotype of the cell, resulting in a rather homogenous phenotype, however this is generally not reflective of what is observed in patient samples, where a hierarchy of cells is maintained. The

expression of a particular mutation could be controlled by a cell specific transcription factor. It is also possible that an abnormality only has the opportunity to occur in a certain cell type. Furthermore, it may be that a mutation is molecularly dependent on a cell-specific co-factor. Research has shown that patients with mutations in *CEBPA* have a distinctive immunophenotype[119] however this was examining the bulk of the disease, not the more primitive compartments.

The lack of engraftment for AMLs containing *CEBPA* mutations, which confer favourable prognosis, is consistant with research presented by our laboratory that the ability of a particular AML to engraft in the NOD/SCID model is related to the prognosis of individual AML cases. AMLs with poor prognosis were able to readily engraft in the NOD/SCID model, while those with favourable prognosis are unable to engraft. Specifically, examination of the follow-up of younger patients with intermediate-risk AML revealed a significant difference in overall survival between NOD/SCID-engrafting and non-engrafting cases. We could not detect a difference between engrafting and non-engrafting cases in various engraftment variables including: homing ability, AML initiating cell frequency, immune rejection by the host or alternative tissue sources. Hence, the ability to engraft NOD/SCID recipients seems to be an inherent property of the cells that is directly related to prognosis.

A possible explanation is that NOD/SCID engraftment reflects the stem cell nature of each individual AML case. AML cases that engraft in the

NOD/SCID assay at six weeks may represent diseases driven by potent leukaemia-initiating cells with stem cell-like self-renewal and proliferation abilities whereas non-engrafting AML cases may involve less potent leukaemia-initiating cells with more restricted progenitor-type self-renewal and proliferation abilities.

In the case of *CEBPA* this ties in with the effects we observed when examining the function of mutations in this gene. While it was clear that the mutations had an effect on early prognitors this same effect was not observed in the *in vivo* assays. Technical difficulties aside, the lack of engraftment in Lin- MNCs by cells transduced with the "N" construct could be explained if the N-terminal mutation offers no advantage over normal HSCs. This fits with the hypothesis that an effect on progenitors is required, and fits with the theory of engraftment being dependant on the cell type that drives the AML. Furthermore, this indicates that *CEBPA* mutations alone are not enough to result in overt leukaemia but may possibly provide a more suspectable target cell for further genetic events This is supported by long latency periods in familial cases involving germline "N" mutations indicating that an accumulation of events is required for development of AML.

From the work presented in this thesis we are able to conclude that gene specific mutations play an important role in leukaemogenesis for patients with normal cytogenetics. Due to the observation of a high proportion of leukaemias containing mutations that disrupt both proliferation and differentiation it is likely that cooperative mutations are required for

leukaemogenesis. Elucidating the order by which mutations arise in AML provides insight into disease progression and the process of leukaemogenesis. Through SNP arrays, we were able to investigate the phenomenon of aUPD and demonstrate that mitotic recombination may be acting as a second hit in leukaemogenesis. Furthermore the examination of areas of aUPD can point to the location of novel genetic aberrations, resulting in our laboratory's further investigations of mutations in the *WT1* gene. The development of mutation specific real-time PCR assays for *CEBPA* offers an additional technique for examination of minimal residual disease in patients with these mutations and has allowed us to examine more closely where the mutations are present within the stem cell and early progenitor compartments of AML. Finally, studying the function of mutations involved in AML and their effects on normal haematopoiesis has provided insight into the target cell for *CEBPA* mutations and what role different mutation combinations may play in AML.

Throughout this thesis, we have demonstrated that there is a wealth of information that can be gained from studying gene specific mutations in AML. Continuing research into these kinds of mutations will one day lead to better understanding of the disease, more accurate prognosis and tailored therapies for patients.

REFERENCES

1. Estey, E. & Dohner, H. Acute myeloid leukaemia. *Lancet* **368**, 1894-907 (2006).
2. Lister, A.R.a.T.A. in Leukemia, 7th edition (ed. Edward S Henderson, T.A.L.a.M.F.G.) (Saunders, 2002).
3. Passegue, E., Jamieson, C.H., Ailles, L.E. & Weissman, I.L. Normal and leukemic hematopoiesis: are leukemias a stem cell disorder or a reacquisition of stem cell characteristics? *Proc Natl Acad Sci U S A* **100 Suppl 1**, 11842-9 (2003).
4. Bonnet, D. Normal and leukaemic stem cells. *Br J Haematol* **130**, 469-79 (2005).
5. Mrozek, K., Marcucci, G., Paschka, P., Whitman, S.P. & Bloomfield, C.D. Clinical relevance of mutations and gene-expression changes in adult acute myeloid leukemia with normal cytogenetics: are we ready for a prognostically prioritized molecular classification? *Blood* **109**, 431-48 (2007).
6. CRUK. (Cancer Research UK, 2006). http://info.cancerresearchuk.org/cancerstats/types/leukaemia/
7. (Office for National Statistics, 2006). http://www.statistics.gov.uk/statbase/Product.asp?vlnk=8843&More=N
8. Hoffbrand, D.C.a.A.V. in Postgraduate Haematology 4th Edition (ed. A.V. Hoffbrand, S.M.L.a.E.D.G.T.) 373-404 (Hodder Arnold, 1998).
9. A.V. Hoffbrand, J.E.P.a.P.A.H.M. Essential Haematology, 4th Edition (Blackwell Science, 2001).
10. Grimwade, D. et al. The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. The Medical Research Council Adult and Children's Leukaemia Working Parties. *Blood* **92**, 2322-33 (1998).
11. Harris, N.L. et al. World Health Organization classification of neoplastic diseases of the hematopoietic and lymphoid tissues: report of the Clinical Advisory Committee meeting-Airlie House, Virginia, November 1997. *J Clin Oncol* **17**, 3835-49 (1999).
12. Appelbaum, F.R., Rowe, J.M., Radich, J. & Dick, J.E. Acute myeloid leukemia. *Hematology Am Soc Hematol Educ Program*, 62-86 (2001).

13. Visani, G. et al. The prognostic value of cytogenetics is reinforced by the kind of induction/consolidation therapy in influencing the outcome of acute myeloid leukemia--analysis of 848 patients. *Leukemia* **15**, 903-9 (2001).
14. Stone, R.M., O'Donnell, M.R. & Sekeres, M.A. Acute myeloid leukemia. *Hematology Am Soc Hematol Educ Program*, 98-117 (2004).
15. Morgan, M.A. & Reuter, C.W. Molecularly targeted therapies in myelodysplastic syndromes and acute myeloid leukemias. *Ann Hematol* **85**, 139-63 (2006).
16. Raanani, P. & Ben-Bassat, I. Detection of minimal residual disease in acute myelogenous leukemia. *Acta Haematol* **112**, 40-54 (2004).
17. Kern, W., Schoch, C., Haferlach, T. & Schnittger, S. Monitoring of minimal residual disease in acute myeloid leukemia. *Crit Rev Oncol Hematol* **56**, 283-309 (2005).
18. Inoue, K. et al. WT1 as a new prognostic factor and a new marker for the detection of minimal residual disease in acute leukemia. *Blood* **84**, 3071-9 (1994).
19. Campana, D. & Coustan-Smith, E. Minimal residual disease studies by flow cytometry in acute leukemia. *Acta Haematol* **112**, 8-15 (2004).
20. Kondo, M. et al. Biology of hematopoietic stem cells and progenitors: implications for clinical application. *Annu Rev Immunol* **21**, 759-806 (2003).
21. Ho, A.D. Kinetics and symmetry of divisions of hematopoietic stem cells. *Exp Hematol* **33**, 1-8 (2005).
22. Breems, D.A., Blokland, E.A., Neben, S. & Ploemacher, R.E. Frequency analysis of human primitive haematopoietic stem cell subsets using a cobblestone area forming cell assay. *Leukemia* **8**, 1095-104 (1994).
23. Ploemacher, R.E., van der Sluijs, J.P., Voerman, J.S. & Brons, N.H. An in vitro limiting-dilution assay of long-term repopulating hematopoietic stem cells in the mouse. *Blood* **74**, 2755-63 (1989).
24. Hao, Q.L., Smogorzewska, E.M., Barsky, L.W. & Crooks, G.M. In vitro identification of single CD34+CD38- cells with both lymphoid and myeloid potential. *Blood* **91**, 4145-51 (1998).
25. Miller, J.S., McCullar, V., Punzel, M., Lemischka, I.R. & Moore, K.A. Single adult human CD34(+)/Lin-/CD38(-) progenitors give rise to natural killer cells, B-lineage cells, dendritic cells, and myeloid cells. *Blood* **93**, 96-106 (1999).

26. Osawa, M. et al. In vivo self-renewal of c-Kit+ Sca-1+ Lin(low/-) hemopoietic stem cells. *J Immunol* **156**, 3207-14 (1996).
27. Spangrude, G.J., Heimfeld, S. & Weissman, I.L. Purification and characterization of mouse hematopoietic stem cells. *Science* **241**, 58-62 (1988).
28. Lapidot, T. et al. Cytokine stimulation of multilineage hematopoiesis from immature human cells engrafted in SCID mice. *Science* **255**, 1137-41 (1992).
29. Shultz, L.D. et al. Multiple defects in innate and adaptive immunologic function in NOD/LtSz-scid mice. *J Immunol* **154**, 180-91 (1995).
30. Cashman, J.D. et al. Kinetic evidence of the regeneration of multilineage hematopoiesis from primitive cells in normal human bone marrow transplanted into immunodeficient mice. *Blood* **89**, 4307-16 (1997).
31. Larochelle, A. et al. Identification of primitive human hematopoietic cells capable of repopulating NOD/SCID mouse bone marrow: implications for gene therapy. *Nat Med* **2**, 1329-37 (1996).
32. Hogan, C.J. et al. Engraftment and development of human CD34(+)-enriched cells from umbilical cord blood in NOD/LtSz-scid/scid mice. *Blood* **90**, 85-96 (1997).
33. Wognum, A.W., Eaves, A.C. & Thomas, T.E. Identification and isolation of hematopoietic stem cells. *Arch Med Res* **34**, 461-75 (2003).
34. Lapidot, T. et al. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* **367**, 645-8 (1994).
35. Bonnet, D. & Dick, J.E. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med* **3**, 730-7 (1997).
36. Singh, S.K. et al. Identification of a cancer stem cell in human brain tumors. *Cancer Res* **63**, 5821-8 (2003).
37. Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J. & Clarke, M.F. Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A* **100**, 3983-8 (2003).
38. O'Brien, C.A., Pollett, A., Gallinger, S. & Dick, J.E. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature* **445**, 106-10 (2007).
39. Guan, Y., Gerhard, B. & Hogge, D.E. Detection, isolation, and stimulation of quiescent primitive leukemic progenitor cells from patients with acute myeloid leukemia (AML). *Blood* **101**, 3142-9 (2003).

40. Cozzio, A. et al. Similar MLL-associated leukemias arising from self-renewing stem cells and short-lived myeloid progenitors. *Genes Dev* **17**, 3029-35 (2003).
41. Hope, K.J., Jin, L. & Dick, J.E. Acute myeloid leukemia originates from a hierarchy of leukemic stem cell classes that differ in self-renewal capacity. *Nat Immunol* **5**, 738-43 (2004).
42. Blair, A., Hogge, D.E., Ailles, L.E., Lansdorp, P.M. & Sutherland, H.J. Lack of expression of Thy-1 (CD90) on acute myeloid leukemia cells with long-term proliferative ability in vitro and in vivo. *Blood* **89**, 3104-12 (1997).
43. So, C.W. et al. MLL-GAS7 transforms multipotent hematopoietic progenitors and induces mixed lineage leukemias in mice. *Cancer Cell* **3**, 161-71 (2003).
44. Rabbitts, T.H. et al. Mouse models of human chromosomal translocations and approaches to cancer therapy. *Blood Cells Mol Dis* **27**, 249-59 (2001).
45. Drynan, L.F. et al. Mll fusions generated by Cre-loxP-mediated de novo translocations can induce lineage reassignment in tumorigenesis. *Embo J* **24**, 3136-46 (2005).
46. Pearce, D.J.a.B., D. (2007).
47. Pearce, D.J. et al. AML engraftment in the NOD/SCID assay reflects the outcome of AML: implications for our understanding of the heterogeneity of AML. *Blood* **107**, 1166-73 (2006).
48. Rombouts, W.J., Martens, A.C. & Ploemacher, R.E. Identification of variables determining the engraftment potential of human acute myeloid leukemia in the immunodeficient NOD/SCID human chimera model. *Leukemia* **14**, 889-97 (2000).
49. Hethcote, H.W. & Knudson, A.G., Jr. Model for the incidence of embryonal cancers: application to retinoblastoma. *Proc Natl Acad Sci U S A* **75**, 2453-7 (1978).
50. Knudson, A.G. Two genetic hits (more or less) to cancer. *Nat Rev Cancer* **1**, 157-62 (2001).
51. Mitelman, F., Johansson, B. & Mertens, F. (February 16, 2007). http://cgap.nci.nih.gov/Chromosomes/Mitelman
52. Nowell, P. & Hungerford, D. A minute chromosome in chronic granulocytic leukemia. *Science* **132** (1960).

53. Kurzrock, R., Kantarjian, H.M., Druker, B.J. & Talpaz, M. Philadelphia chromosome-positive leukemias: from basic mechanisms to molecular therapeutics. *Ann Intern Med* **138**, 819-30 (2003).
54. Pandolfi, P.P. Oncogenes and tumor suppressors in the molecular pathogenesis of acute promyelocytic leukemia. *Hum Mol Genet* **10**, 769-75 (2001).
55. Mitterbauer-Hohendanner, G. & Mannhalter, C. The biological and clinical significance of MLL abnormalities in haematological malignancies. *Eur J Clin Invest* **34 Suppl 2**, 12-24 (2004).
56. So, C.W. & Cleary, M.L. Dimerization: a versatile switch for oncogenesis. *Blood* **104**, 919-22 (2004).
57. Kwok, C., Zeisig, B.B., Dong, S. & So, C.W. Forced homo-oligomerization of RARalpha leads to transformation of primary hematopoietic cells. *Cancer Cell* **9**, 95-108 (2006).
58. Melo, J.V. The molecular biology of chronic myeloid leukaemia. *Leukemia* **10**, 751-6 (1996).
59. Ford, A.M. et al. Fetal origins of the TEL-AML1 fusion gene in identical twins with leukemia. *Proc Natl Acad Sci U S A* **95**, 4584-8 (1998).
60. Small, D. FLT3 Mutations: Biology and Treatment. *Hematology Am Soc Hematol Educ Program*, 178-84 (2006).
61. Stirewalt, D.L. & Radich, J.P. The role of FLT3 in haematopoietic malignancies. *Nat Rev Cancer* **3**, 650-65 (2003).
62. Hiddemann, W. et al. Towards a pathogenesis-oriented therapy of acute myeloid leukemia. *Crit Rev Oncol Hematol* **56**, 235-45 (2005).
63. Chen, W., Rassidakis, G.Z. & Medeiros, L.J. Nucleophosmin gene mutations in acute myeloid leukemia. *Arch Pathol Lab Med* **130**, 1687-92 (2006).
64. Grisendi, S. et al. Role of nucleophosmin in embryonic development and tumorigenesis. *Nature* **437**, 147-53 (2005).
65. Nerlov, C. C/EBPalpha mutations in acute myeloid leukaemias. *Nat Rev Cancer* **4**, 394-400 (2004).
66. Beghini, A. et al. C-kit mutations in core binding factor leukemias. *Blood* **95**, 726-7 (2000).
67. Pabst, T. et al. Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-alpha (C/EBPalpha), in acute myeloid leukemia. *Nat Genet* **27**, 263-70 (2001).

68. Stirewalt, D.L. et al. FLT3, RAS, and TP53 mutations in elderly patients with acute myeloid leukemia. *Blood* **97**, 3589-95 (2001).
69. Preudhomme, C. et al. High incidence of biallelic point mutations in the Runt domain of the AML1/PEBP2 alpha B gene in Mo acute myeloid leukemia and in myeloid malignancies with acquired trisomy 21. *Blood* **96**, 2862-9 (2000).
70. Mrozek, K., Dohner, H. & Bloomfield, C.D. Influence of new molecular prognostic markers in patients with karyotypically normal acute myeloid leukemia: recent advances. *Curr Opin Hematol* **14**, 106-14 (2007).
71. Whitman, S.P. et al. Absence of the wild-type allele predicts poor prognosis in adult de novo acute myeloid leukemia with normal cytogenetics and the internal tandem duplication of FLT3: a cancer and leukemia group B study. *Cancer Res* **61**, 7233-9 (2001).
72. Stirewalt, D.L. et al. Size of FLT3 internal tandem duplication has prognostic significance in patients with acute myeloid leukemia. *Blood* **107**, 3724-6 (2006).
73. Schnittger, S. et al. Nucleophosmin gene mutations are predictors of favorable prognosis in acute myelogenous leukemia with a normal karyotype. *Blood* **106**, 3733-9 (2005).
74. Dohner, K. et al. Mutant nucleophosmin (NPM1) predicts favorable prognosis in younger adults with acute myeloid leukemia and normal cytogenetics: interaction with other gene mutations. *Blood* **106**, 3740-6 (2005).
75. Suzuki, T. et al. Clinical characteristics and prognostic implications of NPM1 mutations in acute myeloid leukemia. *Blood* **106**, 2854-61 (2005).
76. Boissel, N. et al. Prevalence, clinical profile, and prognosis of NPM mutations in AML with normal karyotype. *Blood* **106**, 3618-20 (2005).
77. Frohling, S. et al. CEBPA mutations in younger adults with acute myeloid leukemia and normal cytogenetics: prognostic relevance and analysis of cooperating mutations. *J Clin Oncol* **22**, 624-33 (2004).
78. Bienz, M. et al. Risk assessment in patients with acute myeloid leukemia and a normal karyotype. *Clin Cancer Res* **11**, 1416-24 (2005).
79. Preudhomme, C. et al. Favorable prognostic significance of CEBPA mutations in patients with de novo acute myeloid leukemia: a study from the Acute Leukemia French Association (ALFA). *Blood* **100**, 2717-23 (2002).

80. Barjesteh van Waalwijk van Doorn-Khosrovani, S. et al. Biallelic mutations in the CEBPA gene and low CEBPA expression levels as prognostic markers in intermediate-risk AML. *Hematol J* **4**, 31-40 (2003).
81. Caligiuri, M.A. et al. Rearrangement of ALL1 (MLL) in acute myeloid leukemia with normal cytogenetics. *Cancer Res* **58**, 55-9 (1998).
82. Dohner, K. et al. Prognostic significance of partial tandem duplications of the MLL gene in adult patients 16 to 60 years old with acute myeloid leukemia and normal cytogenetics: a study of the Acute Myeloid Leukemia Study Group Ulm. *J Clin Oncol* **20**, 3254-61 (2002).
83. Marcucci, G. et al. Overexpression of the ETS-related gene, ERG, predicts a worse outcome in acute myeloid leukemia with normal karyotype: a Cancer and Leukemia Group B study. *J Clin Oncol* **23**, 9234-42 (2005).
84. Bergmann, L., Maurer, U. & Weidmann, E. Wilms tumor gene expression in acute myeloid leukemias. *Leuk Lymphoma* **25**, 435-43 (1997).
85. Reilly, J.T. Pathogenesis of acute myeloid leukaemia and inv(16)(p13;q22): a paradigm for understanding leukaemogenesis? *Br J Haematol* **128**, 18-34 (2005).
86. Gilliland, D.G. Hematologic malignancies. *Curr Opin Hematol* **8**, 189-91 (2001).
87. Fenske, T.S. et al. Stem cell expression of the AML1/ETO fusion protein induces a myeloproliferative disorder in mice. *Proc Natl Acad Sci U S A* **101**, 15184-9 (2004).
88. Higuchi, M. et al. Expression of a conditional AML1-ETO oncogene bypasses embryonic lethality and establishes a murine model of human t(8;21) acute myeloid leukemia. *Cancer Cell* **1**, 63-74 (2002).
89. Schessl, C. et al. The AML1-ETO fusion gene and the FLT3 length mutation collaborate in inducing acute leukemia in mice. *J Clin Invest* **115**, 2159-68 (2005).
90. Coyle, T. & Najfeld, V. Translocation (3;21) in Philadelphia chromosome-positive chronic myelogenous leukemia prior to the onset of blast crisis. *Am J Hematol* **27**, 56-9 (1988).
91. Rubin, C.M. et al. Association of a chromosomal 3;21 translocation with the blast phase of chronic myelogenous leukemia. *Blood* **70**, 1338-42 (1987).

92. Pinheiro, R.F. et al. FLT3 mutation and AML/ETO in a case of Myelodysplastic syndrome in transformation corroborates the two hit model of leukemogenesis. *Leuk Res* (2006).
93. Cuenco, G.M. & Ren, R. Both AML1 and EVI1 oncogenic components are required for the cooperation of AML1/MDS1/EVI1 with BCR/ABL in the induction of acute myelogenous leukemia in mice. *Oncogene* **23**, 569-79 (2004).
94. Kelly, L.M. et al. PML/RARalpha and FLT3-ITD induce an APL-like disease in a mouse model. *Proc Natl Acad Sci U S A* **99**, 8283-8 (2002).
95. Le Beau, M.M. et al. Recurring chromosomal abnormalities in leukemia in PML-RARA transgenic mice identify cooperating events and genetic pathways to acute promyelocytic leukemia. *Blood* **102**, 1072-4 (2003).
96. Horwitz, M. The genetics of familial leukemia. *Leukemia* **11**, 1347-59 (1997).
97. Tartaglia, M. et al. Somatic mutations in PTPN11 in juvenile myelomonocytic leukemia, myelodysplastic syndromes and acute myeloid leukemia. *Nat Genet* **34**, 148-50 (2003).
98. Horwitz, M. et al. Genetic heterogeneity in familial acute myelogenous leukemia: evidence for a second locus at chromosome 16q21-23.2. *Am J Hum Genet* **61**, 873-81 (1997).
99. Osato, M., Yanagida, M., Shigesada, K. & Ito, Y. Point mutations of the RUNx1/AML1 gene in sporadic and familial myeloid leukemias. *Int J Hematol* **74**, 245-51 (2001).
100. Smith, M.L., Cavenagh, J.D., Lister, T.A. & Fitzgibbon, J. Mutation of CEBPA in familial acute myeloid leukemia. *N Engl J Med* **351**, 2403-7 (2004).
101. Sellick, G.S. et al. Loci other than 21q22.12 (RUNX1) and 16q21-23.2 cause familial AML. *Leukemia* **19**, 465-6 (2005).
102. Nanri, T. et al. in American Society of Hematology (Orlando, USA, 2006).
103. Orkin, S.H. Diversification of haematopoietic stem cells to specific lineages. *Nat Rev Genet* **1**, 57-64 (2000).
104. Tenen, D.G., Hromas, R., Licht, J.D. & Zhang, D.E. Transcription factors, normal myeloid development, and leukemia. *Blood* **90**, 489-519 (1997).
105. Tenen, D.G. Disruption of differentiation in human cancer: AML shows the way. *Nat Rev Cancer* **3**, 89-101 (2003).

106. Ramji, D.P. & Foka, P. CCAAT/enhancer-binding proteins: structure, function and regulation. *Biochem J* **365**, 561-75 (2002).
107. Frohling, S. & Dohner, H. Disruption of C/EBPalpha function in acute myeloid leukemia. *N Engl J Med* **351**, 2370-2 (2004).
108. Genatlas. (2007). http://www.dsi.univ-paris5.fr/genatlas/fiche.php?n=564
109. Calkhoven, C.F., Muller, C. & Leutz, A. Translational control of C/EBPalpha and C/EBPbeta isoform expression. *Genes Dev* **14**, 1920-32 (2000).
110. Miller, M., Shuman, J.D., Sebastian, T., Dauter, Z. & Johnson, P.F. Structural basis for DNA recognition by the basic region leucine zipper transcription factor CCAAT/enhancer-binding protein alpha. *J Biol Chem* **278**, 15178-84 (2003).
111. Nerlov, C. & Ziff, E.B. CCAAT/enhancer binding protein-alpha amino acid motifs with dual TBP and TFIIB binding ability co-operate to activate transcription in both yeast and mammalian cells. *Embo J* **14**, 4318-28 (1995).
112. Friedman, A.D. & McKnight, S.L. Identification of two polypeptide segments of CCAAT/enhancer-binding protein required for transcriptional activation of the serum albumin gene. *Genes Dev* **4**, 1416-26 (1990).
113. Mueller, B.U. & Pabst, T. C/EBPalpha and the pathophysiology of acute myeloid leukemia. *Curr Opin Hematol* **13**, 7-14 (2006).
114. Leroy, H. et al. CEBPA point mutations in hematological malignancies. *Leukemia* **19**, 329-34 (2005).
115. Schuster, M.B. & Porse, B.T. C/EBPalpha: a tumour suppressor in multiple tissues? *Biochim Biophys Acta* **1766**, 88-103 (2006).
116. Halmos, B. et al. Down-regulation and antiproliferative role of C/EBPalpha in lung cancer. *Cancer Res* **62**, 528-34 (2002).
117. Gery, S. et al. Down-regulation and growth inhibitory role of C/EBPalpha in breast cancer. *Clin Cancer Res* **11**, 3184-90 (2005).
118. Snaddon, J. et al. Mutations of CEBPA in acute myeloid leukemia FAB types M1 and M2. *Genes Chromosomes Cancer* **37**, 72-8 (2003).
119. Lin, L.I. et al. Characterization of CEBPA mutations in acute myeloid leukemia: most patients with CEBPA mutations have biallelic mutations and show a distinct immunophenotype of the leukemic cells. *Clin Cancer Res* **11**, 1372-9 (2005).

120. Akasaka, T. et al. Five members of the CEBP transcription factor family are targeted by recurrent IGH-translocations in B-cell precursor acute lymphoblastic leukemia (BCP-ALL). *Blood* (2006).
121. Pabst, T. et al. AML1-ETO downregulates the granulocytic differentiation factor C/EBPalpha in t(8;21) myeloid leukemia. *Nat Med* **7**, 444-51 (2001).
122. Helbling, D. et al. The leukemic fusion gene AML1-MDS1-EVI1 suppresses CEBPA in acute myeloid leukemia by activation of Calreticulin. *Proc Natl Acad Sci U S A* **101**, 13312-7 (2004).
123. Helbling, D. et al. CBFB-SMMHC is correlated with increased calreticulin expression and suppresses the granulocytic differentiation factor CEBPA in AML with inv(16). *Blood* **106**, 1369-75 (2005).
124. Perrotti, D. et al. BCR-ABL suppresses C/EBPalpha expression through inhibitory action of hnRNP E2. *Nat Genet* **30**, 48-58 (2002).
125. Radomska, H.S. et al. Block of C/EBP alpha function by phosphorylation in acute myeloid leukemia with FLT3 activating mutations. *J Exp Med* **203**, 371-81 (2006).
126. Keeshan, K. et al. Tribbles homolog 2 inactivates C/EBPalpha and causes acute myelogenous leukemia. *Cancer Cell* **10**, 401-11 (2006).
127. Wang, N.D. et al. Impaired energy homeostasis in C/EBP alpha knockout mice. *Science* **269**, 1108-12 (1995).
128. Zhang, D.E. et al. Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice. *Proc Natl Acad Sci U S A* **94**, 569-74 (1997).
129. Flodby, P., Barlow, C., Kylefjord, H., Ahrlund-Richter, L. & Xanthopoulos, K.G. Increased hepatic cell proliferation and lung abnormalities in mice deficient in CCAAT/enhancer binding protein alpha. *J Biol Chem* **271**, 24753-60 (1996).
130. Zhang, P. et al. Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBP alpha. *Immunity* **21**, 853-63 (2004).
131. Heath, V. et al. C/EBPalpha deficiency results in hyperproliferation of hematopoietic progenitor cells and disrupts macrophage development in vitro and in vivo. *Blood* **104**, 1639-47 (2004).
132. Radomska, H.S. et al. CCAAT/enhancer binding protein alpha is a regulatory switch sufficient for induction of granulocytic development from bipotential myeloid progenitors. *Mol Cell Biol* **18**, 4301-14 (1998).

133. Wang, D., D'Costa, J., Civin, C.I. & Friedman, A.D. C/EBPalpha directs monocytic commitment of primary myeloid progenitors. *Blood* **108**, 1223-9 (2006).
134. Cleaves, R., Wang, Q.F. & Friedman, A.D. C/EBPalphap30, a myeloid leukemia oncoprotein, limits G-CSF receptor expression but not terminal granulopoiesis via site-selective inhibition of C/EBP DNA binding. *Oncogene* **23**, 716-25 (2004).
135. Schwieger, M. et al. A dominant-negative mutant of C/EBPalpha, associated with acute myeloid leukemias, inhibits differentiation of myeloid and erythroid progenitors of man but not mouse. *Blood* **103**, 2744-52 (2004).
136. Liu, T.X. et al. Dominant-interfering C/EBPalpha stimulates primitive erythropoiesis in zebrafish. *Exp Hematol* **35**, 230-9 (2007).
137. Sanger, F., Nicklen, S. & Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463-7 (1977).
138. Demaison, C. et al. High-level transduction and gene expression in hematopoietic repopulating cells using a human immunodeficiency [correction of imunodeficiency] virus type 1-based lentiviral vector containing an internal spleen focus forming virus promoter. *Hum Gene Ther* **13**, 803-13 (2002).
139. Zufferey, R., Donello, J.E., Trono, D. & Hope, T.J. Woodchuck hepatitis virus posttranscriptional regulatory element enhances expression of transgenes delivered by retroviral vectors. *J Virol* **73**, 2886-92 (1999).
140. Grimwade, D. The clinical significance of cytogenetic abnormalities in acute myeloid leukaemia. *Best Pract Res Clin Haematol* **14**, 497-529 (2001).
141. Byrd, J.C. et al. Pretreatment cytogenetic abnormalities are predictive of induction success, cumulative incidence of relapse, and overall survival in adult patients with de novo acute myeloid leukemia: results from Cancer and Leukemia Group B (CALGB 8461). *Blood* **100**, 4325-36 (2002).
142. Slovak, M.L. et al. Karyotypic analysis predicts outcome of preremission and postremission therapy in adult acute myeloid leukemia: a Southwest Oncology Group/Eastern Cooperative Oncology Group Study. *Blood* **96**, 4075-83 (2000).
143. Smith, M.L. et al. Development of a human acute myeloid leukaemia screening panel and consequent identification of novel gene mutation in FLT3 and CCND3. *Br J Haematol* **128**, 318-23 (2005).

144. Verhaak, R.G. et al. Mutations in nucleophosmin (NPM1) in acute myeloid leukemia (AML): association with other gene abnormalities and previously established gene expression signatures and their favorable prognostic significance. *Blood* **106**, 3747-54 (2005).
145. Peterson, L.F. et al. The hematopoietic transcription factor AML1 (RUNX1) is negatively regulated by the cell cycle protein cyclin D3. *Mol Cell Biol* **25**, 10205-19 (2005).
146. Bernardin-Fried, F. et al. AML1/RUNX1 increases during G1 to S cell cycle progression independent of cytokine-dependent phosphorylation and induces cyclin D3 gene expression. *J Biol Chem* **279**, 15678-87 (2004).
147. Sicinska, E. et al. Requirement for cyclin D3 in lymphocyte development and T cell leukemias. *Cancer Cell* **4**, 451-61 (2003).
148. Pruneri, G. et al. Cyclin D3 immunoreactivity in follicular lymphoma is independent of the t(6;14)(p21.1;q32.3) translocation or cyclin D3 gene amplification and is correlated with histologic grade and Ki-67 labeling index. *Int J Cancer* **112**, 71-7 (2004).
149. Moller, M.B., Nielsen, O. & Pedersen, N.T. Cyclin D3 expression in non-Hodgkin lymphoma. Correlation with other cell cycle regulators and clinical features. *Am J Clin Pathol* **115**, 404-12 (2001).
150. Paul, J.T. et al. Cyclin D expression in chronic lymphocytic leukemia. *Leuk Lymphoma* **46**, 1275-85 (2005).
151. Filipits, M. et al. Cyclin D3 is a predictive and prognostic factor in diffuse large B-cell lymphoma. *Clin Cancer Res* **8**, 729-33 (2002).
152. Sonoki, T. et al. Cyclin D3 is a target gene of t(6;14)(p21.1;q32.3) of mature B-cell malignancies. *Blood* **98**, 2837-44 (2001).
153. Shaughnessy, J., Jr. et al. Cyclin D3 at 6p21 is dysregulated by recurrent chromosomal translocations to immunoglobulin loci in multiple myeloma. *Blood* **98**, 217-23 (2001).
154. Ebert, M.P. et al. Induction and expression of cyclin D3 in human pancreatic cancer. *J Cancer Res Clin Oncol* **127**, 449-54 (2001).
155. Lopez-Beltran, A. et al. Cyclin D3 expression in primary Ta/T1 bladder cancer. *J Pathol* **209**, 106-13 (2006).
156. Watson, D.S. et al. Cyclin D3 expression, cell proliferation and pathological stage of human primary colorectal cancer. *Oncology* **56**, 66-72 (1999).
157. Roebroek, A.J. et al. The structure of the human c-fes/fps proto-oncogene. *Embo J* **4**, 2897-903 (1985).

158. Hackenmiller, R., Kim, J., Feldman, R.A. & Simon, M.C. Abnormal Stat activation, hematopoietic homeostasis, and innate immunity in c-fes-/- mice. *Immunity* **13**, 397-407 (2000).
159. Lionberger, J.M. & Smithgall, T.E. The c-Fes protein-tyrosine kinase suppresses cytokine-independent outgrowth of myeloid leukemia cells induced by Bcr-Abl. *Cancer Res* **60**, 1097-103 (2000).
160. Bardelli, A. et al. Mutational analysis of the tyrosine kinome in colorectal cancers. *Science* **300**, 949 (2003).
161. Smithgall, T.E., Johnston, J.B., Bustin, M. & Glazer, R.I. Elevated expression of the c-fes proto-oncogene in adult human myeloid leukemia cells in the absence of gene amplification. *J Natl Cancer Inst* **83**, 42-6 (1991).
162. Kim, J. & Feldman, R.A. Activated Fes protein tyrosine kinase induces terminal macrophage differentiation of myeloid progenitors (U937 cells) and activation of the transcription factor PU.1. *Mol Cell Biol* **22**, 1903-18 (2002).
163. Levine, R.L. et al. The JAK2V617F activating mutation occurs in chronic myelomonocytic leukemia and acute myeloid leukemia, but not in acute lymphoblastic leukemia or chronic lymphocytic leukemia. *Blood* **106**, 3377-9 (2005).
164. Baxter, E.J. et al. Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. *Lancet* **365**, 1054-61 (2005).
165. James, C., Ugo, V., Casadevall, N., Constantinescu, S.N. & Vainchenker, W. A JAK2 mutation in myeloproliferative disorders: pathogenesis and therapeutic and scientific prospects. *Trends Mol Med* **11**, 546-54 (2005).
166. Cazzaniga, G. et al. Nucleophosmin mutations in childhood acute myelogenous leukemia with normal karyotype. *Blood* **106**, 1419-22 (2005).
167. Noguera, N.I. et al. Simultaneous detection of NPM1 and FLT3-ITD mutations by capillary electrophoresis in acute myeloid leukemia. *Leukemia* **19**, 1479-82 (2005).
168. Summers, K. et al. Wilms' tumour 1 mutations are associated with FLT3-ITD and failure of standard induction chemotherapy in patients with normal karyotype AML. *Leukemia* **21**, 550-1; author reply 552 (2007).

169. Marcucci, G., Mrozek, K. & Bloomfield, C.D. Molecular heterogeneity and prognostic biomarkers in adults with acute myeloid leukemia and normal cytogenetics. *Curr Opin Hematol* **12**, 68-75 (2005).
170. Bowen, D.T. et al. RAS mutation in acute myeloid leukemia is associated with distinct cytogenetic subgroups but does not influence outcome in patients younger than 60 years. *Blood* **106**, 2113-9 (2005).
171. Ley, T.J. et al. A pilot study of high-throughput, sequence-based mutational profiling of primary human acute myeloid leukemia cell genomes. *Proc Natl Acad Sci U S A* **100**, 14275-80 (2003).
172. Godde, R. et al. Electrophoresis of DNA in human genetic diagnostics - state-of-the-art, alternatives and future prospects. *Electrophoresis* **27**, 939-46 (2006).
173. Hayashi, K. & Yandell, D.W. How sensitive is PCR-SSCP? *Hum Mutat* **2**, 338-46 (1993).
174. O'Donovan, M.C. et al. Blind analysis of denaturing high-performance liquid chromatography as a tool for mutation detection. *Genomics* **52**, 44-9 (1998).
175. Liu, W., Smith, D.I., Rechtzigel, K.J., Thibodeau, S.N. & James, C.D. Denaturing high performance liquid chromatography (DHPLC) used in the detection of germline and somatic mutations. *Nucleic Acids Res* **26**, 1396-400 (1998).
176. Ellis, L.A., Taylor, C.F. & Taylor, G.R. A comparison of fluorescent SSCP and denaturing HPLC for high throughput mutation scanning. *Hum Mutat* **15**, 556-64 (2000).
177. Wenz, H.M., Dailey, D. & Johnson, M.D. Development of a high-throughput capillary electrophoresis protocol for DNA fragment analysis. *Methods Mol Biol* **163**, 3-17 (2001).
178. Li, L. et al. FLT3/ITD expression increases expansion, survival and entry into cell cycle of human haematopoietic stem/progenitor cells. *Br J Haematol* **137**, 64-75 (2007).
179. Scheijen, B. & Griffin, J.D. Tyrosine kinase oncogenes in normal hematopoiesis and hematological disease. *Oncogene* **21**, 3314-33 (2002).
180. Dutt, A. & Beroukhim, R. Single nucleotide polymorphism array analysis of cancer. *Curr Opin Oncol* **19**, 43-9 (2007).
181. Project, H.G. (August 29, 2006). http://www.ornl.gov/sci/techresources/Human_Genome/faq/snps.shtml

182. Kimchi-Sarfaty, C. et al. A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science* **315**, 525-8 (2007).
183. Raghavan, M. et al. Genome-wide single nucleotide polymorphism analysis reveals frequent partial uniparental disomy due to somatic recombination in acute myeloid leukemias. *Cancer Res* **65**, 375-8 (2005).
184. Kottaridis, P.D., Gale, R.E. & Linch, D.C. Prognostic implications of the presence of FLT3 mutations in patients with acute myeloid leukemia. *Leuk Lymphoma* **44**, 905-13 (2003).
185. Chou, W.C. et al. Nucleophosmin mutations in de novo acute myeloid leukemia: the age-dependent incidences and the stability during disease evolution. *Cancer Res* **66**, 3310-6 (2006).
186. Shih, L.Y. et al. AML patients with CEBPalpha mutations mostly retain identical mutant patterns but frequently change in allelic distribution at relapse: a comparative analysis on paired diagnosis and relapse samples. *Leukemia* **20**, 604-9 (2006).
187. Manz, M.G., Miyamoto, T., Akashi, K. & Weissman, I.L. Prospective isolation of human clonogenic common myeloid progenitors. *Proc Natl Acad Sci U S A* **99**, 11872-7 (2002).
188. Miyagawa, K. et al. Mutations of the WT1 gene in childhood nonlymphoid hematological malignancies. *Genes Chromosomes Cancer* **25**, 176-83 (1999).
189. Silva, F.P. et al. Identification of RUNX1/AML1 as a classical tumor suppressor gene. *Oncogene* **22**, 538-47 (2003).
190. Murthy, S.K., DiFrancesco, L.M., Ogilvie, R.T. & Demetrick, D.J. Loss of heterozygosity associated with uniparental disomy in breast carcinoma. *Mod Pathol* **15**, 1241-50 (2002).
191. White, V.A., Chambers, J.D., Courtright, P.D., Chang, W.Y. & Horsman, D.E. Correlation of cytogenetic abnormalities with the outcome of patients with uveal melanoma. *Cancer* **83**, 354-9 (1998).
192. Shearer, P.D. et al. Hemizygous deletions of chromosome band 16q24 in Wilms tumor: detection by fluorescence in situ hybridization. *Cancer Genet Cytogenet* **115**, 100-5 (1999).
193. Hagstrom, S.A. & Dryja, T.P. Mitotic recombination map of 13cen-13q14 derived from an investigation of loss of heterozygosity in retinoblastomas. *Proc Natl Acad Sci U S A* **96**, 2952-7 (1999).

194. Tischfield, J.A. Loss of heterozygosity or: how I learned to stop worrying and love mitotic recombination. *Am J Hum Genet* **61**, 995-9 (1997).
195. Tiesmeier, J. et al. Evidence for allelic evolution of C/EBPalpha mutations in acute myeloid leukaemia. *Br J Haematol* **123**, 413-9 (2003).
196. Pflumio, F. et al. Phenotype and function of human hematopoietic cells engrafting immune-deficient CB17-severe combined immunodeficiency mice and nonobese diabetic-severe combined immunodeficiency mice after transplantation of human cord blood mononuclear cells. *Blood* **88**, 3731-40 (1996).
197. Ogawa, M. Changing phenotypes of hematopoietic stem cells. *Exp Hematol* **30**, 3-6 (2002).
198. Asou, H. et al. Establishment of the acute myeloid leukemia cell line Kasumi-6 from a patient with a dominant-negative mutation in the DNA-binding region of the C/EBPalpha gene. *Genes Chromosomes Cancer* **36**, 167-74 (2003).
199. Suh, H.C. et al. C/EBPalpha determines hematopoietic cell fate in multipotential progenitor cells by inhibiting erythroid differentiation and inducing myeloid differentiation. *Blood* **107**, 4308-16 (2006).
200. Fukuchi, Y. et al. Comprehensive analysis of myeloid lineage conversion using mice expressing an inducible form of C/EBP alpha. *Embo J* **25**, 3398-410 (2006).
201. Cammenga, J. et al. Induction of C/EBPalpha activity alters gene expression and differentiation of human CD34+ cells. *Blood* **101**, 2206-14 (2003).
202. Fitzgibbon, J. et al. Association between acquired uniparental disomy and homozygous gene mutation in acute myeloid leukemias. *Cancer Res* **65**, 9152-4 (2005).
203. Fitzgibbon, J. et al. Genome-wide detection of recurring sites of uniparental disomy in follicular and transformed follicular lymphoma. *Leukemia* **21**, 1514-20 (2007).
204. Sjoblom, T. et al. The consensus coding sequences of human breast and colorectal cancers. *Science* **314**, 268-74 (2006).
205. Steidl , C.S., A. Ebralidze ,H.J. Han ,B. Will ,F. Rosenbauer ,S. Kobayashi ,K.B. O'Brien ,D. Haase ,T. Kohwi-Shigematsu ,D.G. Tenen. in European Society of Haematology (Vienna, 2007).

206. Falini, B., Nicoletti, I., Martelli, M.F. & Mecucci, C. Acute myeloid leukemia carrying cytoplasmic/mutated nucleophosmin (NPMc+ AML): biologic and clinical features. *Blood* **109**, 874-85 (2007).